Mathematical Finance, Vol. 13, No. 2 (April 2003), 245–276

EFFICIENT UNIVERSAL PORTFOLIOS FOR PAST-DEPENDENT TARGET CLASSES

JASON E. CROSS AND ANDREW R. BARRON

Yale University

We present a new universal portfolio algorithm that achieves almost the same level of wealth as could be achieved by knowing stock prices ahead of time. Specifically the algorithm tracks the best in hindsight wealth achievable within target classes of linearly parameterized portfolio sequences. The target classes considered are more general than the standard constant rebalanced portfolio class and permit portfolio sequences to exhibit a continuous form of dependence on past prices or other side information. A primary advantage of the algorithm is that it is easily computable in a polynomial number of steps by way of simple closed-form expressions. This provides an edge over other universal algorithms that require both an exponential number of computations and numerical approximation.

KEY WORDS: universal portfolios, online algorithm, portfolio theory

1. INTRODUCTION

Suppose an investor has a crystal ball that reveals stock prices at any point in the future. Using information gleaned from the ball, the investor selects an investment strategy among a *target class* of portfolio sequences $\mathcal{B}$ to which investment is limited. Assuming rational behavior, the investor chooses the portfolio sequence in $\mathcal{B}$ that maximizes terminal wealth or, equivalently, terminal growth of wealth. Since this optimal growth is achieved by knowing future prices prior to investment it would seem unlikely that optimal growth could be achieved without benefit of the crystal ball.

Surprisingly, this need not be the case and there is a growing literature concerning portfolio algorithms that asymptotically grow wealth at the optimal "crystal ball rate" without the crystal ball. To be more precise, for certain choices of target class $\mathcal{B}$ it is possible to construct portfolio algorithms independent of future prices which, for any given price outcome, asymptotically achieve the same growth rate as the best hindsight strategy in $\mathcal{B}$. The convergence of the growth rate to this optimal hindsight rate is not done merely in a stochastic sense but rather uniformly over all price sequences satisfying certain path properties. For this reason these algorithms are referred to as *universal portfolios* because they achieve optimal growth universally over price paths.

Our aim here is to study universal portfolio strategies that make use of past wealth information or other side information in terms of smoothly parameterized rules, and to seek computationally feasible schemes to achieve close to optimal in hindsight wealth. By way of background we first mention constant rebalanced portfolios in discrete and

Manuscript received August 1999; final revision received March 2002.

Address correspondence to Dr. Jason Cross, 2735 Arthur St. NE, Minneapolis, MN 55418; e-mail: jason.cross@onebox.com.

continuous time. Constant rebalanced portfolios are portfolio sequences that maintain a fixed proportion of wealth in each underlying investment. For example, if an investor starts with half of his wealth in cash and the other half in stock, the constant rebalanced portfolio strategy mandates that the investor buy and sell enough stock at the end of each trading period to return to the initial half/half wealth allocation.

Measuring time in, say, months t and trading in n times per month, the wealth at time t achieved by investing one unit of wealth in a constant rebalanced portfolio $\mathbf{b}$ on m stocks and cash is given by

$$W_t^{(n)}(\mathbf{b}) = \prod_{k=1}^{nt} \mathbf{b}'\mathbf{x}_{k/n},$$

where $\mathbf{x}_{k/n} = (x_{0,k/n}, x_{1,k/n}, \ldots, x_{m,k/n})$ and k/n denotes a fractional time (i.e., $k/n = 3/2$ is halfway through the second month). Here, $x_{0,k/n} = 1$ denotes the wealth relative to cash and $(x_{1,k/n}, \ldots, x_{m,k/n})$ denotes the wealth relatives to stocks over the kth period. As we shall see for stocks satisfying certain basic path properties, for large n this wealth is in close agreement with a corresponding "continuous" wealth,

$$W_t(\mathbf{b}) = \exp\left\{\sum_{j=1}^{m} b_j \mu_{t,j} + \frac{1}{2}\sum_{j=1}^{m} b_j K_{t,j,j} - \frac{1}{2}\sum_{i,j=1}^{m} b_i K_{t,i,j} b_j\right\}, \tag{1.1}$$

where vector $\boldsymbol{\mu}_t$ and matrix $\mathbf{K}_t$ are empirical measures of stock price drift and covariance respectively. This result is in concert with results of Merton (1969, 1971) and Larson (1986) who also showed that the continuous wealth is expressible as the exponential of a quadratic in $\mathbf{b}$. It should be noted that expression (1.1) remains valid for cases where b_i is negative (i.e., when the stock is sold short) and that (1.1) also assumes investment in a riskless asset (in particular, a fraction $b_0 = 1 - \sum_{i=1}^{m} b_i$ is held in cash).

The best in hindsight wealths for both the discrete and continuous cases, $\max_{\mathbf{b}} W_t^{(n)}(\mathbf{b})$ and $\max_{\mathbf{b}} W_t(\mathbf{b})$, naturally depends on the realized path of stock prices. In an effort to track optimal wealth in the discrete case, Cover (1991) introduced the idea of universal portfolios that achieve wealth within a polynomial factor of $\max_{\mathbf{b}} W_t^{(n)}(\mathbf{b})$ without a priori knowledge of the realized price path. The suggested universal portfolio invests in the wealth-weighted average

$$\widehat{\mathbf{b}}_{k/n}^{(n)} = \frac{\int \mathbf{b}\, W_{(k-1)/n}^{(n)}(\mathbf{b})\, d\pi(\mathbf{b})}{\int W_{(k-1)/n}^{(n)}(\mathbf{b})\, d\pi(\mathbf{b})}, \tag{1.2}$$

and achieves wealth

$$\widehat{W}_{k/n}^{(n)} = \int W_{k/n}^{(n)}(\mathbf{b})\, d\pi(\mathbf{b}).$$

In this paper we have two main aims. One is to find universal procedures for target classes beyond the usual constant rebalanced portfolio class. All of the classes to be considered are composed of portfolio sequences $\mathbf{b}(\boldsymbol{\theta}, \mathbf{s}_t)$ that depend in a smoothly parameterized way on the past month's stock information or other side information $\mathbf{s}_t$. In this setting, the class of constant rebalanced portfolios become a special case. Cover and Ordentlich (1996) developed universal procedures in the case where $\mathbf{s}_t$ takes a finite number of values. In addition to the finite state case, our development encompasses more general forms of dependence where it is now permissible for $\mathbf{s}$ to take values in a continuum. Universality with respect to these more flexible target classes is achieved by taking wealth-weighted averages over $\boldsymbol{\theta}$.

Our second aim is to uncover a simple and easily computed universal strategy. Unfortunately, evaluation of the integrals in universal portfolio (1.2) can be problematic, especially when more than a few stocks are considered. Computation requires numerical integration, and computations grow exponentially with the dimension of $\mathbf{b}$ (i.e., the number of stocks). These issues are raised in Singer et al. (1998) where a more efficient universal procedure, computable in a linear number of steps with the number of stocks, is presented. We too seek a more efficient algorithm, although with different methodology and for different target classes of portfolios $\mathcal{B}$. Our approach takes advantage of the continuous time expression (1.1) of wealth $W_t(\mathbf{b})$ to form a portfolio universal with respect to the continuous time target wealth $\max_{\mathbf{b}} W_t(\mathbf{b})$. In particular, for large n we substitute continuous time wealth $W_t(\mathbf{b})$ in place of $W_t^{(n)}(\mathbf{b})$ in expression (1.2) to define a new portfolio

$$\widehat{\mathbf{b}}_{k/n} = \frac{\int \mathbf{b}\, W_{(k-1)/n}(\mathbf{b})\, d\pi(\mathbf{b})}{\int W_{(k-1)/n}(\mathbf{b})\, d\pi(\mathbf{b})}$$

at time $t = (k-1)/n$. Since (1.1) shows the wealth $W_t(\mathbf{b})$ to be an exponential of a quadratic in $\mathbf{b}$, we see that a choice of initial allocation according to a Gaussian π leads to an explicitly and easily computable universal portfolio. This is really just an application of the Bayesian concept of normal conjugation in disguise. Although $\widehat{\mathbf{b}}_{k/n}$ is easy to compute, it remains a continuously traded portfolio. To be truly useful we would like to find a discrete analog of $\widehat{\mathbf{b}}_{k/n}$ that would remain universal. The last section of the paper shows how we can achieve this goal by replacing wealth $W_{(k-1)/n}(\mathbf{b})$ in the derivation of $\widehat{\mathbf{b}}_{k/n}$ with a discretized approximation. This produces a discretely traded, easily computed, universal portfolio that has an elegant closed-form representation.

The resulting portfolio will be computable in m^2 time (m the number of stocks) which is an improvement over portfolio (1.2) computable exponentially in m. Moreover, this computationally advantageous universal portfolio extends to the case of generalized target classes with parameterized portfolio mappings $\mathbf{b}(\boldsymbol{\theta}, \mathbf{s}_t)$ having linear dependence on $\boldsymbol{\theta}$. The result is an efficiently computable universal procedure that can be adapted to a wide variety of target classes and target wealths.

As a final note, we claimed earlier that the expression of wealth (1.1) of the continuously traded portfolios would be the crux of our analysis in later sections of the paper. Since (1.1) holds for cases of short selling (i.e., $b_i < 0$), we will typically assume short selling is permissible in these later sections. However for earlier discussions on discrete time investment we limit considerations to the usual long positions (i.e., $b_i \geq 0$).

Lastly, since (1.1) depends on being able to invest in a riskless asset, we will assume throughout the paper that a portion of assets, b_0, is held in cash. Despite this convention, the discrete time results of Section 3 do not explicitly require an investment in cash. Rather, the convention is maintained here solely to provide continuity with the rest of the paper.

2. PRELIMINARIES

In order to motivate our main results we begin with a general discussion of universal portfolios in a discrete time setting. We begin with the challenge of allocating wealth among m stocks and cash at the start of n consecutive trading periods indexed by $i \in \{1, \ldots, n\}$. The allocation at the beginning of period i can be represented by a portfolio vector $\mathbf{b}_i = (b_{i,0}, b_{i,1}, \ldots, b_{i,m})$ lying in the simplex $B^+ = \left\{\mathbf{b}_i : \sum_{j=0}^{m} b_{i,j} = 1 \text{ and } b_{i,j} \geq 0\right\}$, where $b_{i,0}$ represents the proportion of total wealth in cash and $b_{i,j}$ represents the proportion in stock j at the start of period i.

Letting $p_{i,j}$ be the price of stock j at the end of period i, we define the *wealth relative*, $x_{i,j} = p_{i,j}/p_{i-1,j}$, to be the ratio of the stock price at the end of period i to that at the beginning of period i. Collectively we write the vector of wealth relatives as

$$\mathbf{x}_i = (x_{i,0}, x_{i,1}, \ldots, x_{i,m}),$$

with the understanding that the first component $x_{i,0}$, the wealth relative of cash,[1] is 1. Given these definitions we see that an investor starting with initial wealth W_0 and investing in the sequence of portfolios $\mathbf{b}_1, \ldots, \mathbf{b}_n$ yields a wealth after n periods of

$$W_n = W_0 \prod_{i=1}^{n} \mathbf{b}_i' \mathbf{x}_i.$$

Imagine that an investor is given a crystal ball or some other oracle that would allow knowledge of wealth relatives $\mathbf{x}_i$ up to n periods in the future. Given this information, the best strategy would be to put all wealth in the best stock each period. To set a more modest target, suppose the oracle restricts attention to portfolio sequences $\mathbf{b}_1, \ldots, \mathbf{b}_n$ within some set of sequences $\mathcal{B}$ (the class of constant rebalanced portfolios is one possibility although we will eventually consider more general classes of parameterized sequences). Therefore, given a sequence of wealth relative vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ belonging to sequence class $\mathcal{X}$, the best strategy would be to choose the optimal portfolio sequence in $\mathcal{B}$, call it $\{\mathbf{b}_i^*\}_{i=1}^n$, that maximizes wealth or, equivalently, the empirical growth rate of wealth,

$$R_n \equiv \frac{1}{n} \log W_n / W_0.$$

We label this optimal wealth W_n^* and the associated optimal growth rate R_n^*.

Suppose now that such an oracle is lacking. Is it still possible to achieve the optimal W_n^* without knowing $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ beforehand? Intuition suggests not and in a strict sense this is usually the case. However for some choices of $\mathcal{B}$ we can do the next best thing, which is to achieve the optimal growth rate of wealth R_n^*. Specifically we can sometimes construct portfolio sequences $\{\widehat{\mathbf{b}}_i\}_{i=1}^n$ independent of future knowledge of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ that have growth rate $\widehat{R}_n$ asymptotically equal to R_n^*. Essentially, such a strategy will achieve without future knowledge of prices almost the same wealth as if we were given future knowledge of prices and then allowed to act according to any portfolio sequence in $\mathcal{B}$. We call such a portfolio sequence $\{\widehat{\mathbf{b}}_i\}_{i=1}^n$ a *universal portfolio* because it achieves the optimal growth rate uniformly over all price paths in $\mathcal{X}$.

DEFINITION 2.1. Let R_n^* be the maximal growth rate achievable over a set of portfolio sequences $\mathcal{B}$ when given future knowledge of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ lying in a set of permissible wealth relative sequences $\mathcal{X}$. We call a portfolio sequence $\{\widehat{\mathbf{b}}_i\}_{i=1}^n$ determined independently of future knowledge a *universal portfolio* (or simply *universal*) with respect to $\mathcal{B}$ and $\mathcal{X}$ if its corresponding growth rate $\widehat{R}_n$ is such that

$$\lim_{n\to\infty} \sup_{\{\mathbf{x}_i\}_{i=1}^n \in \mathcal{X}} \left(R_n^* - \widehat{R}_n\right) = \lim_{n\to\infty} \sup_{\{\mathbf{x}_i\}_{i=1}^n \in \mathcal{X}} \frac{1}{n} \log \frac{W_n^*}{\widehat{W}_n} \leq 0.$$

[1] Here one might think of cash as a proxy for any riskless asset. For simplicity we assume it has a zero rate of return although we can easily extend subsequent analysis to any nonzero risk-free rate of return r by replacing wealth relative $X_{i,j}$ with the discounted wealth relative $Y_{i,j} = (1+r)^{-1} X_{i,j}$. Under this substitution expressions of wealth are now discounted to their net present value at time 0. However the fundamental conclusions of the paper remain unaffected.

Here $W_n^* = W_0 \prod_{i=1}^n \mathbf{b}_i^{*\prime}\mathbf{x}_i$ and $\widehat{W}_n = W_0 \prod_{i=1}^n \widehat{\mathbf{b}}_i^{\prime}\mathbf{x}_i$ are the wealths of the best sequence in $\mathcal{B}$ chosen in hindsight and the universal portfolio respectively. Additionally, we refer to $\mathcal{B}$ as the *target class* of the universal portfolio, R_n^* as the *target growth rate*, and $W_n^* = W_0 \exp\{n R_n^*\}$ as the *target wealth*.

We should note that to successfully consider target classes $\mathcal{B}$ beyond the class of constant rebalanced portfolios requires us to use a more restrictive class of wealth relative sequences $\mathcal{X}$ than the set of all nonnegative wealth relative sequences previously considered by Cover (1991). In general we limit consideration to subsets of sequences satisfying certain regularity conditions, one being that the wealth relatives are bounded above zero. That said, when limiting consideration to only constant rebalanced portfolios, for the discrete time results of the next section one may allow for all stock price sequences, as Cover has shown.

3. A UNIVERSAL PORTFOLIO IN DISCRETE TIME

3.1. Introduction and Analysis

To set the foundation for future results we first present a simple universal portfolio algorithm traded in discrete time. The portfolio is universal for certain parameterized target classes $\mathcal{B}$ having possible dependence on various types of side information (past prices, economic indicators, expert opinion, etc.). Information from time i is summarized through a state variable $\mathbf{s}_i$ which takes values in some arbitrary domain S and is used to form portfolios at the start of the subsequent investment period.

To define a parameterized target class we consider a parameter space $\Theta \subseteq \mathbf{R}^d$ and portfolio map $\mathbf{b} : \Theta \times S \to B^+$ that for each period i sets the portfolio $\mathbf{b}_i \equiv \mathbf{b}(\boldsymbol{\theta}, \mathbf{s}_{i-1})$. We think of the class as a set of functions $\{\mathbf{b}(\boldsymbol{\theta}, \cdot) : \boldsymbol{\theta} \in \Theta\}$. Each function in the class, or equivalently each $\boldsymbol{\theta} \in \Theta$, defines a distinct sequence of portfolios whose allocations are determined at the start of period i through $\mathbf{b}(\boldsymbol{\theta}, \mathbf{s}_{i-1})$.

There are many different types of portfolio sequence classes that can be modeled in this framework. For instance we have already briefly discussed the class of constant rebalanced portfolios in the introduction. A constant rebalanced portfolio is a sequence of portfolios for which the same portfolio is used at the start of each trading period. Thus if we begin by using portfolio $\mathbf{b}$ we buy and sell enough of each stock at the end of each trading period so that our wealth proportions return to $\mathbf{b}$ by the start of the next trading period. For investment in cash and m stocks one possible parameterization of this class is given by the parameter space $\Theta = \left\{\boldsymbol{\theta} \in \mathbf{R}^m : \sum_{j=1}^m \theta_j \leq 1, \theta_j \geq 0\right\}$ and mapping $\mathbf{b}_i \equiv \mathbf{b}(\boldsymbol{\theta}) = \left(1 - \sum_{j=1}^m \theta_j, \theta_1, \ldots, \theta_m\right)$. In this case the state variable $\mathbf{s}_i$ is left undefined because constant rebalanced portfolios do not use side information.

For an example of a class that uses past prices in forming the portfolio, suppose an investor follows an investment strategy where he allocates wealth according to the most recent stock gains. If he believes the outperforming stocks will continue to outperform in the next period he might consider investment in a portfolio such as $\mathbf{s}_i = \mathbf{x}_i / \sum_{j=0}^m x_{i,j}$. However the weak form of the efficient market hypothesis suggests such a trend should not exist, so the investor hedges his bets by splitting wealth between $\mathbf{x}_i / \sum_{j=0}^m x_{i,j}$ and a constant rebalanced portfolio. Consider the class of portfolio sequence for which, before investment, the investor fixes a constant rebalanced portfolio and a fraction of wealth to put between the constant rebalanced portfolio and most recent wealth relative vector. In this case, our side information is the normalized wealth relative vector

$\mathbf{s}_i = \mathbf{x}_i / \sum_{j=0}^{m} x_{i,j}$, and our parameters $\boldsymbol{\theta}$ are used to set the constant rebalanced portfolio and the fraction of wealth. Hence a possible parameterization of the class would be given by parameter space $\Theta = \{\boldsymbol{\theta} \in \mathbf{R}^{m+1} : 0 \le \theta_j \le 1, \sum_{j=1}^{m} \theta_j \le 1\}$ and mapping $\mathbf{b}_i \equiv \mathbf{b}(\boldsymbol{\theta}, \mathbf{s}_{i-1}) = (1 - \theta_{m+1})\mathbf{s}_{i-1} + \theta_{m+1}(1 - \sum_{j=1}^{m} \theta_j, \theta_1, \ldots, \theta_m)$.

Obviously there are many different possible target classes that we could consider. Regardless, for any given target class $(\Theta, S, \mathbf{b}(\boldsymbol{\theta}, \mathbf{s}))$, sequence of wealth relatives $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, and sequence of side information states $\mathbf{s}_0, \ldots, \mathbf{s}_{n-1}$, the wealth achieved by the portfolio sequence indexed by $\boldsymbol{\theta}$ up to time n is given by

$$W_n(\boldsymbol{\theta}) = W_0 \prod_{i=1}^{n} \mathbf{b}'(\boldsymbol{\theta}, \mathbf{s}_{i-1})\mathbf{x}_i.$$

The optimal wealth (or target wealth) achievable for a particular price outcome is denoted by $W_n^* \equiv \max_{\boldsymbol{\theta} \in \Theta} W_n(\boldsymbol{\theta})$. It is possible that the target wealth could be achieved by more than one portfolio sequence in the target class. Regardless of this, we use $\boldsymbol{\theta}_n^*$ to refer to a particular indexation of an optimizing sequence as of time n, and write $W_n^* = W_n(\boldsymbol{\theta}_n^*)$.

We now endeavor to construct universal portfolios with respect to these parameterized target classes. To this end we consider the portfolio sequence defined by the updating rule

$$\widehat{\mathbf{b}}_i = \frac{\int_\Theta \mathbf{b}(\boldsymbol{\theta}, \mathbf{s}_{i-1}) W_{i-1}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}{\int_\Theta W_{i-1}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}, \tag{3.1}$$

where π is some measure on Θ with $\pi(\Theta) = 1$. At each step, $\widehat{\mathbf{b}}_i$ constitutes a weighted average of portfolios in the target class weighted according to how well these portfolios have done in the past. In essence it is a simple extension of the procedure presented in Cover (1991). Indeed if the target class is set to the class of constant rebalanced portfolios, $\widehat{\mathbf{b}}_i$ becomes the π-weighted version of Cover's universal portfolio.

We now denote the wealth achieved by $\widehat{\mathbf{b}}_i$ up to time n by $\widehat{W}_n = W_0 \prod_{i=1}^{n} \widehat{\mathbf{b}}_i' \mathbf{x}_i$. Note the following lemma.

LEMMA 3.1.

$$\widehat{W}_n = \int_\Theta W_n(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta}).$$

Proof. The proof follows immediately from a telescoping product argument. Note that

$$\begin{aligned}
\widehat{W}_n &= W_0 \prod_{i=1}^{n} \widehat{\mathbf{b}}_i \mathbf{x}_i \\
&= W_0 \prod_{i=1}^{n} \frac{\int_\Theta \mathbf{b}'(\boldsymbol{\theta}, \mathbf{s}_{i-1}) W_{i-1}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}{\int_\Theta W_{i-1}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})} \mathbf{x}_i \\
&= W_0 \prod_{i=1}^{n} \frac{\int_\Theta W_i(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}{\int_\Theta W_{i-1}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})} \\
&= \int_\Theta W_n(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})
\end{aligned}$$ □

The portfolio sequence, $\widehat{\mathbf{b}}_i$, can be interpreted as an implementation of a strategy where we split the initial wealth over a continuum of investment managers, each of whom uses a unique portfolio sequence in the target class indexed by $\boldsymbol{\theta}$. To see this suppose that at the start of each period each manager invests according to his own $\boldsymbol{\theta}$ in portfolio $\mathbf{b}(\boldsymbol{\theta}, \mathbf{s}_{i-1})$. At the end of each period each manager has wealth proportional to $W_i(\boldsymbol{\theta})$. If originally we had split our initial wealth according to π, we would have a collective wealth at time n of $\widehat{W}_n = \int_\Theta W_n(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})$, but as we showed in Lemma 3.1 this is the same wealth achieved by portfolio sequence $\widehat{\mathbf{b}}_i$. So in essence the use of $\widehat{\mathbf{b}}_i$ is equivalent to distributing initial wealth among a continuum of investment managers.

It can be shown that $\widehat{\mathbf{b}}_i$ is universal in the sense that it achieves the same growth rate of wealth as the best in hindsight strategy over the target class $(\Theta, S, \mathbf{b}(\boldsymbol{\theta}, \mathbf{s}))$. Recall that the best in hindsight strategy, indexed by optimal parameter $\boldsymbol{\theta}_n^*$, depends on knowing the particular outcome of wealth relatives $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. At the start of period i the strategy uses the portfolio $\mathbf{b}'(\boldsymbol{\theta}_n^*, \mathbf{s}_{i-1})$ to achieve optimal wealth

$$W_n^* \equiv \max_{\boldsymbol{\theta}\in\Theta} W_n(\boldsymbol{\theta}) = W_0 \prod_{i=1}^n \mathbf{b}'(\boldsymbol{\theta}_n^*, \mathbf{s}_{i-1})\mathbf{x}_i$$

and optimal growth rate of wealth $R_n^* \equiv \frac{1}{n} \log W_n^*/W_0$. In order for $\widehat{\mathbf{b}}_i$ to be universal (i.e., to achieve optimal growth of wealth R_n^* without a priori knowledge of wealth relatives) we must assume the following minimal conditions on target classes and price sequences.

3.1.1. Investment Conditions.

1. There exists constant $L_{\mathbf{x}} > 0$ such that $1/L_{\mathbf{x}} \le x_{i,j} \le L_{\mathbf{x}}$ for all $1 \le i \le n$ and $0 \le j \le m$.
2. $\Theta \subseteq \mathbf{R}^d$ is convex, compact and has positive Lebesgue measure with respect to $\mathbf{R}^d$.
3. π has density $f(\boldsymbol{\theta})$ on Θ (with respect to Lebesgue measure) which is bounded above 0 by some $\delta > 0$.
4. $\mathbf{b}(\boldsymbol{\theta}, \mathbf{s})$ is Lipschitz in that there exists constant $L_{\mathbf{b}} > 0$ independent of $\boldsymbol{\theta}$ and $\mathbf{s}$ such that $\|\mathbf{b}(\boldsymbol{\theta}_0, \mathbf{s}) - \mathbf{b}(\boldsymbol{\theta}_1, \mathbf{s})\| \le L_{\mathbf{b}} \|\boldsymbol{\theta}_0 - \boldsymbol{\theta}_1\|$ for all $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1 \in \Theta$ and $\mathbf{s} \in S$.

With these assumptions we now prove that $\widehat{\mathbf{b}}_i$ is universal.

THEOREM 3.1. *Suppose the investment conditions hold. Then the sequence of portfolios* $\widehat{\mathbf{b}}_i$ *of equation (3.1) is universal with respect to the target class* $(\Theta, S, \mathbf{b}(\boldsymbol{\theta}, \mathbf{s}))$ *and achieves target wealth* $W_n(\boldsymbol{\theta}_n^*) = \max_{\boldsymbol{\theta}\in\Theta} W_n(\boldsymbol{\theta})$ *in the sense that*

$$\frac{W_n(\boldsymbol{\theta}_n^*)}{\widehat{W}_n} = O(n^d)$$

uniformly over all wealth relative sequences $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *satisfying investment condition 1 and possible state sequences* $\mathbf{s}_1, \ldots, \mathbf{s}_n$.

Obviously the above order bound implies $\widehat{\mathbf{b}}_i$ is universal because

$$\begin{aligned}\lim_{n\to\infty} \sup_{\{\mathbf{x}_i\}_{i=1}^n} \left(R_n^* - \widehat{R}_n\right) &= \lim_{n\to\infty} \sup_{\{\mathbf{x}_i\}_{i=1}^n} \frac{1}{n} \log \frac{W_n(\boldsymbol{\theta}_n^*)}{\widehat{W}_n} \\ &= \lim_{n\to\infty} O\left(\frac{\log n}{n}\right) = 0.\end{aligned}$$

Proof. The idea of the proof is to show that the wealth generated by the universal portfolio is concentrated in a small neighborhood of points around $\boldsymbol{\theta}_n^*$ shrinking on order $O(n^{-d})$. For this purpose define the neighborhood $\Theta_n^* = \{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\theta} - \boldsymbol{\theta}_n^*\| \leq 1/n\}$. By our investment assumptions the wealth ratio is bounded on these neighborhoods for all n since

$$\begin{aligned}
\max_{\boldsymbol{\theta}\in\Theta_n^*}\left(\frac{W_n\left(\boldsymbol{\theta}_n^*\right)}{W_n(\boldsymbol{\theta})}\right) &= \max_{\boldsymbol{\theta}\in\Theta_n^*}\left(\frac{W_n\left(\boldsymbol{\theta}_n^*\right) - W_n(\boldsymbol{\theta})}{W_n(\boldsymbol{\theta})} + 1\right) \\
&= \max_{\boldsymbol{\theta}\in\Theta_n^*}\left(\frac{W_0\prod_{i=1}^n \mathbf{x}_i'\left(\mathbf{b}\left(\boldsymbol{\theta}_n^*, \mathbf{s}_{i-1}\right) - \mathbf{b}(\boldsymbol{\theta}, \mathbf{s}_{i-1})\right)}{W_0\prod_{i=1}^n \mathbf{x}_i'\mathbf{b}(\boldsymbol{\theta}, \mathbf{s}_{i-1})} + 1\right) \\
&\leq \max_{\boldsymbol{\theta}\in\Theta_n^*}\left(\frac{\prod_{i=1}^n L_{\mathbf{x}} L_{\mathbf{b}}\left\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}\right\|}{\prod_{i=1}^n L_{\mathbf{x}}^{-1}}\right) + 1 \\
&\leq \left(\frac{\prod_{i=1}^n L_{\mathbf{x}} L_{\mathbf{b}}(1/n)}{\prod_{i=1}^n L_{\mathbf{x}}^{-1}}\right) + 1 \\
&\leq \left(\frac{L_{\mathbf{x}}^2 L_{\mathbf{b}}}{n}\right)^n + 1 \\
&< C \quad \text{for some constant } C.
\end{aligned}$$

Now use the bound on the wealth ratio along with Lemma to note that

$$\begin{aligned}
\frac{\widehat{W}_n}{W_n^*} &= \int_{\Theta} \frac{W_n(\boldsymbol{\theta})}{W_n^*}\, d\pi(\boldsymbol{\theta}) \\
&\geq \int_{\Theta_n^*} \frac{W_n(\boldsymbol{\theta})}{W_n^*}\, d\pi(\boldsymbol{\theta}) \\
&= (1/C)\pi(\Theta_n^*).
\end{aligned}$$

The rest of the proof hinges on bounding the π-measure of Θ_n^*. Since the volume of Θ_n^* decreases as $O(n^{-d})$ and because we assume that π's derivative with respect to Lebesgue measure is strictly positive it follows that there exists a constant $R > 0$ such that $\pi(\Theta_n^*) \geq Rn^{-d}$. Therefore,

$$\frac{\widehat{W}_n}{W_n^*} \geq (1/C)Rn^{-d}. \tag{3.2}$$

Inverting the ratio we conclude that $\frac{W_n^*}{\widehat{W}_n}$ is $O(n^d)$.

3.2. Computational Issues

Although portfolio $\widehat{\mathbf{b}}_i$ of (3.1) has the desirable property of universality, its practical usage is limited by its computational properties. Unfortunately for any choice of measure π the calculation of (3.1) generally requires computations that grow exponentially with the dimensionality of Θ, as for parameter spaces of more than a few dimensions calculation can become prohibitively intensive. For most choices of target class and weighting measure π, the portfolio $\widehat{\mathbf{b}}_i$ lacks a simple formulaic expression and must be computed via numerical integration.

For these reasons we are motivated to find similar procedures that are explicitly computable at a faster polynomial rate. Our approach will be to use continuous time versions of the procedures we have developed thus far. With the specific choice of a Gaussian π and $\mathbf{b}(\boldsymbol{\theta}, \mathbf{s})$ linear in $\boldsymbol{\theta}$ we will find that our proposed procedure has a simple closed-form expression that can be computed on order $\max\{m^2, d^2\}$ steps (m being the number of stocks and d being the dimensionality of the parameter space Θ) at any time instance t.

4. UNIVERSALITY IN CONTINUOUS TIME

4.1. Continuously Traded Target Classes

Let us now consider investment in a sequence of continuously traded portfolios selected among m stocks and cash. As before, we denote a portfolio by the vector $\mathbf{b} = (b_0, \widetilde{\mathbf{b}}) = (b_0, b_1, \dots, b_m)$, where vector $\widetilde{\mathbf{b}} = (b_1, \dots, b_m)$ holds the proportions of wealth put in each stock and b_0 denotes the proportion of wealth put in cash. To ensure that the portfolio is self financing we again require that $\sum_{j=0}^{m} b_j = 1$. However, unlike the discrete case, we no longer assume that the b_j are nonnegative. In other words we now allow for short selling and purchase on margin. For convenience we henceforth refer to the set of all such portfolios as B, and write

$$B = \left\{ \mathbf{b} \in R^m : \sum_{j=0}^{m} b_j = 1 \right\}.$$

We wish to consider continuous time trading within each of T discrete time periods indexed by $\tau \in \{1, \dots, T\}$. As before, we assume that side information is used at the start of each period to determine the portfolio to be used for the rest of the period. Again we assume that such side information is represented by some variable $\mathbf{s}$ taking values in domain S. The idea will be to use the side information to select a constant rebalanced portfolio at the start of each period and then continuously trade that portfolio for the rest of the period.

The classes of portfolio sequences we wish to consider are rigorously defined through a portfolio mapping $\mathbf{b} : \mathbf{R}^d \times S \to B$. Note that unlike in the previous section the parameter space of the mapping is now the entire $\mathbf{R}^d$ space rather than an arbitrary subspace Θ. Also unlike the previous section we choose to restrict attention to those portfolio maps that are linear in $\boldsymbol{\theta}$. Specifically, we consider mappings $\mathbf{b}(\boldsymbol{\theta}, \mathbf{s}) = (b_0(\boldsymbol{\theta}, \mathbf{s}), \widetilde{\mathbf{b}}(\boldsymbol{\theta}, \mathbf{s}))$ with cash component

$$b_0(\boldsymbol{\theta}, \mathbf{s}) = 1 - \sum_{j=1}^{m} b_j(\boldsymbol{\theta}, \mathbf{s})$$

and vector of stock components

$$\widetilde{\mathbf{b}}(\boldsymbol{\theta}, \mathbf{s}) = (b_1(\boldsymbol{\theta}, \mathbf{s}), \dots, b_m(\boldsymbol{\theta}, \mathbf{s})) = \mathbf{A}(\mathbf{s})\boldsymbol{\theta}.$$

Here $\mathbf{A}(\mathbf{s})$ is an $m \times d$ linear transformation dependent on the state of side information $\mathbf{s}$. Thus we write

$$\mathbf{b}(\boldsymbol{\theta}, \mathbf{s}) = (b_0(\boldsymbol{\theta}, \mathbf{s}), \widetilde{\mathbf{b}}(\boldsymbol{\theta}, \mathbf{s})) = \left(1 - \sum_{j=1}^{m} \mathbf{A}_j(\mathbf{s}) \cdot \boldsymbol{\theta}, \mathbf{A}(\mathbf{s})\boldsymbol{\theta} \right), \tag{4.1}$$

where $\mathbf{A}_j(\mathbf{s})$ denotes the jth row of matrix $\mathbf{A}(\mathbf{s})$.

It should be noted that our restriction to linear portfolio maps allows us to take advantage of a special property that we will derive shortly. Specifically, we shall see that the wealth generated by continuously trading these linear portfolios has a convenient closed-form expression. In turn, this expression drives the rest of the paper by allowing us to construct easily computable universal portfolios. In general, similar closed-form expressions of wealth do not exist for nonlinear mappings and hence we impose this restriction.

To proceed, the types of strategies we wish to consider are as follows. At the start of period 1, use the available side information $\mathbf{s}_0$ and portfolio mapping $\mathbf{b}(\boldsymbol{\theta}, \mathbf{s})$ to set the constant rebalanced portfolio $\mathbf{b}_1 = \mathbf{b}(\boldsymbol{\theta}, \mathbf{s}_0)$. Then take $\mathbf{b}_1$ and trade it continuously over time period $t \in (0, 1]$. At the start of the second period, take side information $\mathbf{s}_1$ and set the constant rebalanced portfolio $\mathbf{b}_2 = \mathbf{b}(\boldsymbol{\theta}, \mathbf{s}_1)$ and trade it continuously over period $t \in (1, 2]$. Repeat the process T times until the invest horizon is reached at time $t = T$.

As before, we think of the *target class* as a triplet of parameter space, side information domain, and portfolio map (i.e., $(\mathbf{R}^d, S, \mathbf{b}(\boldsymbol{\theta}, \mathbf{s}))$) or, equivalently for the linear mapping case, as a triplet of parameter space, side information domain, and linear transformation $(\mathbf{R}^d, S, \mathbf{A}(\mathbf{s}))$. Every $\boldsymbol{\theta} \in \mathbf{R}^d$ corresponds to a member of the class that represents a sequence of T constant rebalanced portfolios traded continuously over T time periods.

We now wish to calculate the wealth achieved by these strategies. This may be done by looking at the wealth achieved by rebalancing only n times a period and then taking the limit as $n \to \infty$. Thus let $p_{t,j}$ denote the price of stock j at time $t \in [0, T]$. As before, we suppose that we start with some initial wealth W_0. Clearly the end wealth achieved by strategy $\boldsymbol{\theta}$ rebalanced n times a period is

$$
\begin{aligned}
(4.2) \qquad W_T^{(n)}(\boldsymbol{\theta}) &= W_0 \prod_{\tau=1}^{T} \prod_{k=1}^{n} \left(b_0(\boldsymbol{\theta}, \mathbf{s}_{\tau-1}) + \sum_{j=1}^{m} b_j(\boldsymbol{\theta}, \mathbf{s}_{\tau-1}) \frac{p_{k\tau/n,j}}{p_{(k-1)\tau/n,j}} \right) \\
&= W_0 \prod_{\tau=1}^{T} \prod_{k=1}^{n} \left(1 + \sum_{j=1}^{m} b_j(\boldsymbol{\theta}, \mathbf{s}_{\tau-1}) \left(\frac{p_{k\tau/n,j}}{p_{(k-1)\tau/n,j}} - 1 \right) \right).
\end{aligned}
$$

We would like to know what happens as $n \to \infty$ (i.e., when we trade continuously). Thus we seek an expression for the limiting wealth $W_T(\boldsymbol{\theta}) = \lim_{n\to\infty} W_T^{(n)}(\boldsymbol{\theta})$ for an arbitrary price path $p_{t,j}$. To derive this limit it will be convenient to work with the log price path $Z_t \equiv (\log p_{t,1}, \ldots, \log p_{t,m})$. For any realization of this path we define the empirical log-drift over period τ as

$$\boldsymbol{\mu}_\tau \equiv (\mu_{\tau,1}, \ldots, \mu_{\tau,m}) = (Z_{\tau,1} - Z_{\tau-1,1}, \ldots, Z_{\tau,m} - Z_{\tau-1,m}).$$

Similarly, we also define the sequence of empirical covariance matrices $\mathbf{K}_\tau^{(n)}$ for period τ having entries

$$K_{\tau,i,j}^{(n)} \equiv \sum_{k=1}^{n} \left(Z_{(k/n)\tau,i} - Z_{((k-1)/n)\tau,i} \right) \left(Z_{(k/n)\tau,j} - Z_{((k-1)/n)\tau,j} \right).$$

In order to derive a limiting wealth $W_t(\boldsymbol{\theta})$ we need to assume that price paths are not too wild in their fluctuation. Specifically they need to exhibit finite quadratic variation. We would also like the empirical covariance matrices to converge to some positive definite limit. Henceforth we require the following conditions.

4.1.1. Minimal Path Conditions.

1. There exists a constant $L_p > 0$ dependent on the path such that

$$\left(1 + \frac{L_p(1+\log n)}{\sqrt{n}}\right)^{-1} \leq \frac{p_{k\tau/n,j}}{p_{(k-1)\tau/n,j}} \leq 1 + \frac{L_p(1+\log n)}{\sqrt{n}}$$

for all $n, k \in \{1, \ldots, n\}$, $\tau \in \{1, \ldots, T\}$, and $j \in \{1, \ldots, m\}$.

2. There exists a positive definite matrix $\mathbf{K}_\tau$ such that $\lim_{n\to\infty} \mathbf{K}_\tau^{(n)} = \mathbf{K}_\tau$ elementwise.

Given these assumptions we now derive the wealth achieved by continuously trading $\mathbf{b}(\boldsymbol{\theta}, \mathbf{s}_{\tau-1})$. It should come as no surprise that the expression we derive is in agreement with results previously published by Merton (1969, 1971) and Larson (1986). However, unlike their results, which are proved using an underlying diffusion process for p_t, we choose to stay away from stochastic assumptions and instead presume only path properties. This nonstochastic setting is consistent with our goal of developing universal procedures that have growth optimal properties independent of stochastics.

THEOREM 4.1. *If the minimal path conditions hold, the wealth factor from continuously trading* $\mathbf{b}_\tau = \mathbf{b}(\boldsymbol{\theta}, \mathbf{s}_{\tau-1})$ *each time period is*

$$W_\tau(\boldsymbol{\theta})/W_{\tau-1}(\boldsymbol{\theta}) = \exp\left\{\boldsymbol{\mu}_\tau' \widetilde{\mathbf{b}}_\tau + \frac{1}{2} diag(\mathbf{K}_\tau)' \widetilde{\mathbf{b}}_\tau - \frac{1}{2} \widetilde{\mathbf{b}}_\tau' \mathbf{K}_\tau \widetilde{\mathbf{b}}_\tau\right\}$$

with $\widetilde{\mathbf{b}}_\tau = \mathbf{A}(\mathbf{s}_{\tau-1})\boldsymbol{\theta}$ *and hence the wealth achieved by time T is*

$$W_T(\boldsymbol{\theta}) = W_0 \exp\left\{\sum_{\tau=1}^{T} \boldsymbol{\mu}_\tau' \mathbf{A}(\mathbf{s}_{\tau-1})\boldsymbol{\theta} + \frac{1}{2}\sum_{\tau=1}^{T} diag(\mathbf{K}_\tau)' \mathbf{A}(\mathbf{s}_{\tau-1})\boldsymbol{\theta} - \frac{1}{2}\sum_{\tau=1}^{T} \boldsymbol{\theta}' \mathbf{A}'(\mathbf{s}_{\tau-1})\mathbf{K}_\tau \mathbf{A}(\mathbf{s}_{\tau-1})\boldsymbol{\theta}\right\}, \tag{4.3}$$

an exponential quadratic in $\boldsymbol{\theta}$. *Furthermore, if the* $d \times d$ *matrix*

$$\boldsymbol{\Omega}_T \equiv \mathbf{A}'(\mathbf{s}_0)\mathbf{K}_1\mathbf{A}(\mathbf{s}_0) + \cdots + \mathbf{A}'(\mathbf{s}_{T-1})\mathbf{K}_T\mathbf{A}(\mathbf{s}_{T-1})$$

is invertible, then

$$W_T(\boldsymbol{\theta}) = \exp\left\{-\frac{1}{2}(\boldsymbol{\theta} - \mathbf{v}_T)'\boldsymbol{\Omega}_T(\boldsymbol{\theta} - \mathbf{v}_T) + \frac{1}{2}\mathbf{v}_T'\boldsymbol{\Omega}_T\mathbf{v}_T\right\}, \tag{4.4}$$

where

$$\mathbf{v}_T = \boldsymbol{\Omega}_T^{-1}\left(\mathbf{A}'(\mathbf{s}_0)\left[\boldsymbol{\mu}_1 + \frac{1}{2} diag\ (\mathbf{K}_1)\right] + \cdots + \mathbf{A}'(\mathbf{s}_{T-1})\left[\boldsymbol{\mu}_1 + \frac{1}{2} diag\ (\mathbf{K}_T)\right]\right).$$

Proof. The theorem is a corollary of Lemma 5.1 presented in Section 5.1. Both are proved in the Appendix. □

Although Theorem 4.1 is stated in terms of the terminal wealth $W_T(\boldsymbol{\theta})$ at an integer time T, it is a straightforward exercise to show that the wealth $W_t(\boldsymbol{\theta})$ for any real-valued time $t < T$ is

$$W_t(\boldsymbol{\theta}) = W_0 \exp\left\{ \sum_{\tau=I(t)} (\boldsymbol{\mu}_\tau)' \mathbf{A}(\mathbf{s}_\tau)\boldsymbol{\theta} + \frac{1}{2} \sum_{\tau=I(t)} \text{diag}\,(\mathbf{K}_\tau)' \mathbf{A}(\mathbf{s}_\tau)\boldsymbol{\theta} - \frac{1}{2} \sum_{\tau=I(t)} \boldsymbol{\theta}' \mathbf{A}'(\mathbf{s}_\tau) \mathbf{K}_\tau \mathbf{A}(\mathbf{s}_\tau)\boldsymbol{\theta} \right\}, \tag{4.5}$$

where $I(t) = \{1, \ldots, \lceil t \rceil - 1, t\}$ and $\mathbf{s}_t = \mathbf{s}_{\lceil t \rceil}$. Here $\boldsymbol{\mu}_t$ and $\mathbf{K}_t$ represent the empirical drift and covariance measures over the truncated interval $(\lceil t \rceil - 1, t]$. Similarly, defining

$$\boldsymbol{\Omega}_t \equiv \mathbf{A}'(\mathbf{s}_0)\mathbf{K}_1\mathbf{A}(\mathbf{s}_0) + \cdots + \mathbf{A}'\left(\mathbf{s}_{\lceil t \rceil - 2}\right) \mathbf{K}_{\lceil t \rceil - 1} \mathbf{A}\left(\mathbf{s}_{\lceil t \rceil - 2}\right) + \mathbf{A}'\left(\mathbf{s}_{\lceil t \rceil - 1}\right) \mathbf{K}_t \mathbf{A}\left(\mathbf{s}_{\lceil t \rceil - 1}\right)$$

and

$$\mathbf{v}_t = \boldsymbol{\Omega}_t^{-1} \left(\mathbf{A}'(\mathbf{s}_1) \left[\boldsymbol{\mu}_1 + \frac{1}{2} \text{diag}\,(\mathbf{K}_1) \right] + \cdots + \mathbf{A}'\left(\mathbf{s}_{\lceil t \rceil}\right) \left[\boldsymbol{\mu}_t + \frac{1}{2} \text{diag}\,(\mathbf{K}_t) \right] \right),$$

we can also write $W_t(\boldsymbol{\theta})$ as

$$W_t(\boldsymbol{\theta}) = \exp\left\{ -\frac{1}{2} (\boldsymbol{\theta} - \mathbf{v}_t)' \boldsymbol{\Omega}_t (\boldsymbol{\theta} - \mathbf{v}_t) + \frac{1}{2} \mathbf{v}_t' \boldsymbol{\Omega}_t \mathbf{v}_t \right\}. \tag{4.6}$$

4.2. Continuously Traded Universal Procedures

Our goal now is to find a universal procedure for the continuously traded target classes just described. Many of our results will be similar to those derived by Jamshidian (1992) for the constant rebalanced portfolio class.

In the continuous time setting we think of a universal portfolio as a nonanticipating sequence of portfolios $\{\widehat{\mathbf{b}}_t\}_{t=0}^T$ that generate wealth $\widehat{W}_T$ matching the hindsight optimal wealth $W_T^* \equiv \max_{\boldsymbol{\theta} \in \mathbf{R}^d} W_T(\boldsymbol{\theta})$ to first order in the exponent. An intuitive way to generate such a procedure is to take the discrete-time universal portfolio (3.1) and trade it on finer and finer time scales. Thus, at the start of period $t = k/n$ we set the portfolio

$$\widehat{\mathbf{b}}_t^{(n)} = \frac{\int_\Theta \mathbf{b}\left(\boldsymbol{\theta}, \mathbf{s}_{\lceil t \rceil - 1}\right) W_t^{(n)}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}{\int_\Theta W_t^{(n)}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}.$$

Here we take Θ to be a large subset of $\mathbf{R}^d$ and time increments $1/n$ to be sufficiently small such that $W_t^{(n)}(\boldsymbol{\theta})$ mimics $W_t(\boldsymbol{\theta})$. In the limit as $n \to \infty$ and Θ increases to $\mathbf{R}^d$ this results in continuously trading the portfolio

$$\widehat{\mathbf{b}}_t = \frac{\int_{\mathbf{R}^d} \mathbf{b}\left(\boldsymbol{\theta}, \mathbf{s}_{\lceil t \rceil - 1}\right) W_{t-1/n}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}{\int_{\mathbf{R}^d} W_{t-1/n}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}, \tag{4.7}$$

at each time instance t.

In order to gain simple computation of $\widehat{\mathbf{b}}_t$ it would help to choose a measure $\pi(\boldsymbol{\theta})$ that yielded convenient closed-form expressions for (4.7). Here the Bayesian concept of a normal conjugate prior proves to be very useful. Recall that when a prior Gaussian density is used in conjunction with a Gaussian sampling density the resulting posterior density is also Gaussian. To apply this to the present case note from (4.6) that $W_t(\boldsymbol{\theta})$ is an exponential quadratic in $\boldsymbol{\theta}$ and is hence equivalent to a nonnormalized Gaussian density. Choosing π to be an arbitrary Gaussian measure we conclude from the normal conjugation property that the measure

$$\frac{W_t(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}{\int_{\mathbf{R}^d} W_t(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})} \tag{4.8}$$

is also Gaussian. Therefore upon examining (4.7) we see that the calculation of $\widehat{\mathbf{b}}_t$ is equivalent to a normal expectation calculation.

Suppose then that we choose $\pi \sim N(\boldsymbol{\lambda}, \boldsymbol{\Lambda})$ with $\boldsymbol{\lambda} \in \mathbf{R}^d$ and positive definite $\boldsymbol{\Lambda}$. By completing the square in the exponent of $W_t(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})$, it is quickly verified that (4.8) is Gaussian with mean

$$\psi_t = \left(\boldsymbol{\Omega}_t + \boldsymbol{\Lambda}^{-1}\right)^{-1} \left(\mathbf{A}'(\mathbf{s}_0)\left[\boldsymbol{\mu}_1 + \frac{1}{2}\,\mathrm{diag}\,(\mathbf{K}_1)\right] + \cdots \right.$$
$$\left. + \mathbf{A}'\left(\mathbf{s}_{\lceil t\rceil - 1}\right)\left[\boldsymbol{\mu}_t + \frac{1}{2}\,\mathrm{diag}\,(\mathbf{K}_t)\right] + \boldsymbol{\Lambda}^{-1}\boldsymbol{\lambda}\right)$$

and covariance

$$\left(\boldsymbol{\Omega}_t + \boldsymbol{\Lambda}^{-1}\right)^{-1}.$$

Thus it follows that the stock component, $\widehat{\widetilde{\mathbf{b}}}_t$, of $\{(4.7)\}$ is now

$$\widehat{\widetilde{\mathbf{b}}}_t = \frac{\int_{\mathbf{R}^d} \widetilde{\mathbf{b}}\left(\boldsymbol{\theta}, \mathbf{s}_{\lceil t\rceil - 1}\right) W_t(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}{\int_{\mathbf{R}^d} W_t(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})} \tag{4.9}$$
$$= \frac{\int_{\mathbf{R}^d} \mathbf{A}\left(\mathbf{s}_{\lceil t\rceil - 1}\right) \boldsymbol{\theta}\, W_t(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}{\int_{\mathbf{R}^d} W_t(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}$$
$$= \mathbf{A}(\mathbf{s}_{\lceil t\rceil - 1})\, \psi_t$$

with cash component $\widehat{\mathbf{b}}_{t,0} = 1 - \sum_{j=1}^{m} \left(\mathbf{A}\left(\mathbf{s}_{\lceil t\rceil - 1}\right) \psi_t\right)_j$.

Shortly we will prove the universality of $\widehat{\mathbf{b}}_t$ with respect to instances of the general continuously traded target class $(\mathbf{R}^d, S, \mathbf{A}(\mathbf{s}))$. In preparation for the proof we restrict attention to classes and price paths satisfying the following conditions.

4.2.1. Universality Conditions.

1. The minimal path conditions hold.
2. There exists a constant $L_\mu > 0$ independent of T such that $\frac{1}{T}\sum_{\tau=1}^{T} \|\boldsymbol{\mu}_\tau\| < L_\mu$ for all $T \in \mathbf{Z}^+$.
3. There exists a constant $L_{\mathbf{K}} > 0$ independent of T such that $\frac{1}{T}\sum_{\tau=1}^{T} \lambda_{\max}(\mathbf{K}_\tau) < L_{\mathbf{K}}$, where $\lambda_{\max}(\mathbf{K}_\tau)$ denotes the maximum eigenvalue of $\mathbf{K}_\tau$.
4. For any $\mathbf{s} \in S$, the $m \times d$ matrix $\mathbf{A}(\mathbf{s})$ has full rank and there exists a constant $L_{\mathbf{A}} > 0$ independent of $\mathbf{s}$ such that $\lambda_{\max}(\mathbf{A}(\mathbf{s})\mathbf{A}'(\mathbf{s})) \leq L_{\mathbf{A},d}$.
5. There exists some integer β (possibly dependent on price path and side information sequence) such that $\boldsymbol{\Omega}_t$ is invertible for all times $t > \beta$.
6. For $T > \beta$, there exists positive constants L_Ω^- and L_Ω^+ independent of T such that the minimum and maximum eigenvalues of $\boldsymbol{\Omega}_T$ satisfy, $L_\Omega^- T \leq \lambda_{\min}(\boldsymbol{\Omega}_T) \leq \lambda_{\max}(\boldsymbol{\Omega}_t) \leq L_\Omega^+ T$.

The key to the universality proof is the following lemma, which gives the wealth achieved by trading $\widehat{\mathbf{b}}_t$ continuously. Note the similarity to Lemma 3.1 of Section 3.1.

LEMMA 4.1. *Suppose that the universality conditions hold. Let $\widehat{W}_t$ be the wealth achieved by trading $\widehat{\mathbf{b}}_t$ (with stock component (4.9)) continuously over time interval* $[0, T]$ *with respect to the class* $(\mathbf{R}^d, S, \mathbf{A}(\mathbf{s}))$. *Then*

$$\widehat{W}_T = \int_{\mathbf{R}^d} W_T(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta}) \tag{4.10}$$

for all T.

Proof. The proof follows from building $\widehat{\mathbf{b}}_t$ from discrete time procedures and then examining the wealths of these discrete procedures in the limit.

First consider $\mathbf{b}_\tau = \mathbf{b}(\boldsymbol{\theta}, \mathbf{s}_\tau)$ traded only n times per period for each period $\tau = 1, 2, \ldots, T$. Let $W_T^{(n)}(\boldsymbol{\theta})$ be the wealth achieved by this strategy. We note from Theorem 4.1 that $\lim_{n\to\infty} W_T^{(n)}(\boldsymbol{\theta}) = W_T(\boldsymbol{\theta})$ with $W_T(\boldsymbol{\theta})$ given in (4.3). From Lemma 5.1 (stated and proved in the next section) it can be quickly verified that, for any $\lambda \in \mathbf{R}$ and compact set $\Theta_\lambda = \{\boldsymbol{\theta} \in \mathbf{R}^d : \|\boldsymbol{\theta}\| \le \lambda\}$,

$$\lim_{n\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_\lambda} \left| W_T^{(n)}(\boldsymbol{\theta}) - W_T(\boldsymbol{\theta}) \right| = 0.$$

From this it follows that there exists an N such that $W_T^{(n)}(\boldsymbol{\theta})$ is uniformly bounded over $\boldsymbol{\theta} \in \Theta_\lambda$ and $n \ge N$. It is straightforward to extend this conclusion to $W_t^{(n)}(\boldsymbol{\theta})$ for any fixed real time $t < T$.

Now consider investment in the discrete procedure in which integration over $\mathbf{R}^d$ is replaced by integration over Θ_λ,

$$\widehat{\mathbf{b}}_{t,\lambda}^{(n)} = \frac{\int_{\Theta_\lambda} \mathbf{b}\left(\boldsymbol{\theta}, \mathbf{s}_{\lceil t\rceil - 1}\right) W_t^{(n)}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}{\int_{\Theta_\lambda} W_t^{(n)}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}.$$

The procedure rebalances n times a period for T periods (i.e., we use $\widehat{\mathbf{b}}_{0,\lambda}^{(n)}$ at time $t = 0$, then rebalance according to $\widehat{\mathbf{b}}_{1/n,\lambda}^{(n)}$ at time $t = 1/n$, rebalance to $\widehat{\mathbf{b}}_{2/n,\lambda}^{(n)}$ at $t = 2/n$, etc.). By Lemma 3.1 this strategy yields a wealth of $\widehat{W}_{T,\lambda}^{(n)} = \int_{\Theta_\lambda} W_T^{(n)}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})$. Now increase the size of n. Since the functions $W_t^{(n)}(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta_\lambda$ are bounded uniformly in n, for any given $t \le T$ we can use dominated convergence to show that

$$\widehat{\mathbf{b}}_{t,\lambda} \equiv \lim_{n\to\infty} \widehat{\mathbf{b}}_{t,\lambda}^{(n)} = \frac{\int_{\Theta_\lambda} \mathbf{b}\left(\boldsymbol{\theta}, \mathbf{s}_{\lceil t\rceil - 1}\right) W_t(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}{\int_{\Theta_\lambda} W_t(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}. \tag{4.11}$$

Similarly we can apply dominated convergence once more to prove that

$$\widehat{W}_{t,\lambda} = \lim_{n\to\infty} \widehat{W}_{t,\lambda}^{(n)} = \int_{\Theta_\lambda} W_t(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta}). \tag{4.12}$$

The final step in the proof is to let $\lambda \to \infty$ and allow Θ_λ to increase to all of $\mathbf{R}^d$. Looking at (4.5) it can be seen that $W_t(\boldsymbol{\theta})$ is π-integrable over $\mathbf{R}^d$ when π is taken to be a Gaussian measure. Hence, by one more application of dominated convergence,

$$\widehat{\mathbf{b}}_t \equiv \lim_{\lambda\to\infty} \widehat{\mathbf{b}}_{t,\lambda} = \frac{\int_{\mathbf{R}^d} \mathbf{b}\left(\boldsymbol{\theta}, \mathbf{s}_{\lceil t\rceil - 1}\right) W_t(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}{\int_{\mathbf{R}^d} W_t(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}$$

and

$$\widehat{W}_t \equiv \lim_{\lambda\to\infty} \widehat{W}_{t,\lambda} = \int_{\mathbf{R}^d} W_t(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta}). \qquad \square$$

We now use Lemma 4.1 to prove the universality of $\widehat{\mathbf{b}}_t$. Recall that $\widehat{\mathbf{b}}_t$ is universal with respect to a target class if it comes within a polynomial factor of the hindsight optimal wealth $W_t^* = \max_{\boldsymbol{\theta}\in\mathbf{R}^d} W_t(\boldsymbol{\theta})$.

THEOREM 4.2. *Suppose that the universality conditions hold. Then* $\widehat{\mathbf{b}}_t$ *with stock component (4.9) is universal with respect to the linear target class* $(\mathbf{R}^d, S, \mathbf{A}(\mathbf{s}))$ *in the sense that there is a constant C such that for any* $T > \max\{\beta, \lambda_{\max}(\mathbf{\Lambda}^{-1})/L_{\mathbf{\Omega}}^-\}$

$$\frac{W_T^*}{\widehat{W}_T} \leq CT^{d/2}.$$

Proof. Since $T > \beta$ it follows that $\mathbf{\Omega}_t$ is invertible and expression (4.4) is valid for $W_T(\boldsymbol{\theta})$. Recall by Lemma 4.1 that $\widehat{W}_T = \int_{\mathbf{R}^d} W_T(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})$. Use (4.4) and the fact that $\pi \sim N(\boldsymbol{\lambda}, \mathbf{\Lambda})$ to compute $\int_{\mathbf{R}^d} W_T(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})$ and conclude that

$$\widehat{W}_T = W_0 \left(|\mathbf{\Lambda}| \left|\mathbf{\Omega}_T + \mathbf{\Lambda}^{-1}\right|\right)^{-1/2} \exp\left\{\frac{1}{2}\psi_T' \left(\mathbf{\Omega}_T + \mathbf{\Lambda}^{-1}\right)\psi_T - \frac{1}{2}\boldsymbol{\lambda}'\mathbf{\Lambda}^{-1}\boldsymbol{\lambda}\right\}.$$

Compute W_T^* by maximizing (4.4) over $\boldsymbol{\theta}$. This yields $\boldsymbol{\theta}^* = \mathbf{v}_T$ and $W_T^* = W_0 \exp\left\{\frac{1}{2}\mathbf{v}_T'\mathbf{\Omega}_T\mathbf{v}_T\right\}$. Thus,

$$\frac{W_T^*}{\widehat{W}_T} = \left(|\mathbf{\Lambda}| \left|\mathbf{\Omega}_T + \mathbf{\Lambda}^{-1}\right|\right)^{1/2} \times \exp\left\{-\frac{1}{2}\psi_T' \left(\mathbf{\Omega}_T + \mathbf{\Lambda}^{-1}\right)\psi_T + \frac{1}{2}\mathbf{v}_T'\mathbf{\Omega}_T^{-1}\mathbf{v}_T + \frac{1}{2}\boldsymbol{\lambda}'\mathbf{\Lambda}^{-1}\boldsymbol{\lambda}\right\}. \tag{4.13}$$

We bound each part of the expression in turn. First note that

$$\left(|\mathbf{\Lambda}| \left|\mathbf{\Omega}_T + \mathbf{\Lambda}^{-1}\right|\right)^{1/2} = \left(|\mathbf{\Lambda}\Omega_T + I|\right)^{1/2}.$$

Since $\det \mathbf{A} \leq (\lambda_{\max}(\mathbf{A}))^d$ for any $d \times d$ matrix $\mathbf{A}$, it follows that

$$\left(|\mathbf{\Lambda}\Omega_T + I|\right)^{1/2} \leq \left[\lambda_{\max}\left(\mathbf{\Lambda}\Omega_T + I\right)\right]^{d/2}.$$

But $\lambda_{\max}(\mathbf{\Lambda}\Omega_T + I) \leq \lambda_{\max}(\mathbf{\Lambda})\lambda_{\max}(\mathbf{\Omega}_T) + 1 \leq \lambda_{\max}(\mathbf{\Lambda})L_{\mathbf{\Omega}}^+ T + 1$, so

$$\left(|\mathbf{\Lambda}| \left|\mathbf{\Omega}_T + \mathbf{\Lambda}^{-1}\right|\right)^{1/2} \leq \left[\lambda_{\max}(\mathbf{\Lambda})L_{\mathbf{\Omega}}^+ T + 1\right]^{d/2} = O(T^{d/2}).$$

All that remains is to show the exponential term of (4.13) is of constant order. First note that

(4.14)

$$\begin{aligned}
\|\mathbf{v}_T\| &= \lambda_{\max}\left(\mathbf{\Omega}_T^{-1}\right) \sum_{\tau=1}^{T} \left(\left[\boldsymbol{\mu}_\tau + \frac{1}{2}\operatorname{diag}\mathbf{K}_\tau\right]' \mathbf{A}(\mathbf{s}_{\tau-1})\mathbf{A}'(\mathbf{s}_{\tau-1}) \left[\boldsymbol{\mu}_t + \frac{1}{2}\operatorname{diag}\mathbf{K}_\tau\right]\right)^{1/2} \\
&\leq \frac{1}{\lambda_{\min}(\mathbf{\Omega}_T)} \sum_{\tau=1}^{T} \lambda_{\max}\left(\mathbf{A}\left(\mathbf{s}_{\tau-1}\right)\mathbf{A}'\left(\mathbf{s}_{\tau-1}\right)\right)^{1/2} \left\|\boldsymbol{\mu}_\tau + \frac{1}{2}\operatorname{diag}\mathbf{K}_\tau\right\| \\
&\leq \frac{L_{\mathbf{A},d}^{1/2}}{L_{\mathbf{\Omega}}^- T} \sum_{\tau=1}^{T} \left(\|\boldsymbol{\mu}_\tau\| + \frac{\sqrt{m}}{2}\boldsymbol{\lambda}_{\max}\left(\mathbf{K}_\tau\right)\right) \\
&\leq \frac{L_{\mathbf{A},d}^{1/2}}{L_{\mathbf{\Omega}}^-} L_\mu \frac{\sqrt{m}}{2} L_{\mathbf{K}} = O(1).
\end{aligned}$$

Since $\psi_T = (\mathbf{\Omega}_T + \mathbf{\Lambda}^{-1})^{-1}(\mathbf{\Omega}_T \mathbf{v}_T + \mathbf{\Lambda}^{-1}\boldsymbol{\lambda})$, it follows that

$$\begin{aligned} \|\psi_T\| &= \| \left(\mathbf{\Omega}_T + \Lambda^{-1}\right)^{-1} \left(\mathbf{\Omega}_T \mathbf{v}_T + \Lambda^{-1}\boldsymbol{\lambda}\right) \| \\ &\leq \frac{1}{\lambda_{\min}\left(\mathbf{\Omega}_T + \mathbf{\Lambda}^{-1}\right)} \left\| \mathbf{\Omega}_T \mathbf{v}_T + \mathbf{\Lambda}^{-1}\boldsymbol{\lambda} \right\| \\ &\leq \frac{\lambda_{\max}(\mathbf{\Omega}_T)\|\mathbf{v}_T\| + \|\mathbf{\Lambda}^{-1}\boldsymbol{\lambda}\|}{\lambda_{\min}(\mathbf{\Omega}_T)} \\ &\leq \frac{L_\Omega^+ T O(1) + \|\Lambda^{-1}\boldsymbol{\lambda}\|}{L_\Omega^- T} = O(1). \end{aligned} \tag{4.15}$$

Now take the exponential term of (4.13). Note that

$$\begin{aligned} &\exp\left\{-\frac{1}{2}\psi_T'\left(\mathbf{\Omega}_T + \mathbf{\Lambda}^{-1}\right)\psi_T + \frac{1}{2}\mathbf{v}_T'\mathbf{\Omega}_T\mathbf{v}_T + \frac{1}{2}\boldsymbol{\lambda}'\mathbf{\Lambda}^{-1}\boldsymbol{\lambda}\right\} \\ &= \exp\left\{-\frac{1}{2}(\psi_T + \mathbf{v}_T)'\,\mathbf{\Omega}_T(\psi_T - \mathbf{v}_T) - \frac{1}{2}\psi_T'\mathbf{\Lambda}^{-1}\psi_T + \frac{1}{2}\boldsymbol{\lambda}'\mathbf{\Lambda}^{-1}\boldsymbol{\lambda}\right\} \end{aligned} \tag{4.16}$$

The third term in the exponent is constant by definition and the second term is bounded by virtue of (4.15). Thus the proof is complete if we can show that

$$-\frac{1}{2}(\psi_T + \mathbf{v}_T)'\,\mathbf{\Omega}_T(\psi_T - \mathbf{v}_T)$$

is bounded. Note that

$$\begin{aligned} &-\frac{1}{2}(\psi_T + \mathbf{v}_T)'\,\mathbf{\Omega}_T(\psi_T - \mathbf{v}_T) \\ &\leq \frac{\lambda_{\max}(\mathbf{\Omega}_T)}{2}\|\psi_T + \mathbf{v}_T\|\,\|\psi_T - \mathbf{v}_T\| \\ &\leq \frac{L_\Omega^+ T}{2}\|\psi_T + \mathbf{v}_T\|\left\|\left(\mathbf{\Omega}_T + \mathbf{\Lambda}^{-1}\right)^{-1}\mathbf{\Lambda}^{-1}(\boldsymbol{\lambda} - \mathbf{v}_T)\right\| \\ &\leq \frac{L_\Omega^+ T}{2}\|\psi_T + \mathbf{v}_T\|\frac{\lambda_{\max}\left(\mathbf{\Lambda}^{-1}\right)}{\lambda_{\min}\left(\mathbf{\Omega}_T + \mathbf{\Lambda}^{-1}\right)}\|(\boldsymbol{\lambda} - \mathbf{v}_T)\| \\ &\leq \frac{L_\Omega^+ T}{2}\frac{\lambda_{\max}\left(\mathbf{\Lambda}^{-1}\right)}{\lambda_{\min}(\mathbf{\Omega}_T)}(\|\psi_T\| + \|\mathbf{v}_T\|)(\|\boldsymbol{\lambda}\| + \|\mathbf{v}_T\|) \\ &\leq \frac{L_\Omega^+\lambda_{\max}\left(\mathbf{\Lambda}^{-1}\right)}{2L_\Omega^-}(\|\psi_T\| + \|\mathbf{v}_T\|)(\|\boldsymbol{\lambda}\| + \|\mathbf{v}_T\|) \\ &= O(1) \end{aligned}$$

by virtue of (4.14) and (4.15).

Hence the theorem is proved. □

5. AN EFFICIENT UNIVERSAL PORTFOLIO

5.1. Definition and Main Theorem

We know from Theorem 4.2 that the continuously traded portfolio,

$$\widehat{\mathbf{b}}_t = \left(1 - \sum_{j=1}^{m} \left(\mathbf{A}\left(\mathbf{s}_{\lceil t \rceil - 1}\right) \psi_t\right)_j, \mathbf{A}\left(\mathbf{s}_{\lceil t \rceil - 1}\right) \psi_t\right) \tag{5.1}$$

of Section 4.2 is universal with respect to the continuously traded target class $(\mathbf{R}^d, S, \mathbf{A}(\mathbf{s}))$. Moreover, due to the closed-form expression of ψ_t this portfolio is computable at any time t in a polynomial number of steps. Although the properties of $\widehat{\mathbf{b}}_t$ are certainly desirable, its continuous updating makes it unusable for real-world application. For practicality we would like to find a discretely traded adaptation of $\widehat{\mathbf{b}}_t$ that retains the universality and ease of computation. The thought is to trade portfolio $\widehat{\mathbf{b}}_t$ following formula (5.1) only n times a period (i.e., at times $\left\{\frac{1}{n}, \frac{2}{n}, \ldots, T\right\}$) with nonanticipating discrete-time approximations to ψ_t. Specifically, at time $t = k/n$ we consider trading the portfolio

$$\widehat{\mathbf{b}}_{k/n}^{(n)} \equiv \left(1 - \sum_{j=1}^{m} \left(\mathbf{A}\left(\mathbf{s}_{\lceil k/n \rceil - 1}\right) \psi_{(k-1)/n}^{(n)}\right)_j \mathbf{A}\left(\mathbf{s}_{\lceil k/n \rceil - 1}\right) \psi_{k-1/n}^{(n)}\right), \tag{5.2}$$

where intuitively

$$\begin{aligned} \psi_{k/n}^{(n)} \equiv \left(\boldsymbol{\Omega}_{k/n}^{(n)} + \boldsymbol{\Lambda}^{-1}\right)^{-1} \Bigg(\mathbf{A}'(\mathbf{s}_0)\left[\boldsymbol{\mu}_1 + \frac{1}{2}\operatorname{diag}\left(\mathbf{K}_1^{(n)}\right)\right] + \cdots \\ + \mathbf{A}'\left(\mathbf{s}_{\lceil k/n \rceil - 1}\right)\left[\boldsymbol{\mu}_{k/n} + \frac{1}{2}\operatorname{diag}\left(\mathbf{K}_{k/n}^{(n)}\right)\right]\Bigg) + \boldsymbol{\Lambda}^{-1}\boldsymbol{\lambda}, \end{aligned} \tag{5.3}$$

and

$$\begin{aligned} \boldsymbol{\Omega}_{k/n}^{(n)} &\equiv \mathbf{A}'(\mathbf{s}_0)\mathbf{K}_1^{(n)}\mathbf{A}(\mathbf{s}_0) + \cdots + \mathbf{A}'\left(\mathbf{s}_{\lceil k/n \rceil - 2}\right)\mathbf{K}_{\lceil k/n \rceil - 1}^{(n)}\mathbf{A}\left(\mathbf{s}_{\lceil k/n \rceil - 2}\right) \\ &\quad + \mathbf{A}'\left(\mathbf{s}_{\lceil k/n \rceil - 1}\right)\mathbf{K}_{k/n}^{(n)}\mathbf{A}\left(\mathbf{s}_{\lceil k/n \rceil - 1}\right). \end{aligned}$$

If n is large there is little difference between the wealth achieved by $\widehat{\mathbf{b}}_{k/n}^{(n)}$ and its continuously traded counterpart $\widehat{\mathbf{b}}_t$. Indeed the following analysis culminating in Theorem 5.1 shows this rigorously. The intuition behind the theorem can be summarized in a few steps. First, note that for large n there is little difference between the wealths achieved by trading $\widehat{\mathbf{b}}_t$ continuously and trading it n times per period through formula

$$\widehat{\mathbf{b}}_{k/n} = \frac{\int_{\mathbf{R}^d} \mathbf{b}\left(\boldsymbol{\theta}, \mathbf{s}_{\lceil t \rceil - 1}\right) W_{(k-1)/n}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}{\int_{\mathbf{R}^d} W_{(k-1)/n}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}.$$

However, for large n, $W_{(k-1)/n}(\boldsymbol{\theta})$ is close to the exponential quadratic $W_0 \exp\{\mathbf{q}_{(k-1)/n}(\boldsymbol{\theta})\}$, where

$$\begin{aligned} \mathbf{q}_{k/n}(\boldsymbol{\theta}) \equiv \sum_{\tau \in I(k/n)} \boldsymbol{\mu}_\tau' \mathbf{A}\left(\mathbf{s}_{\lceil t \rceil - 1}\right)\boldsymbol{\theta} + \frac{1}{2} \sum_{\tau \in I(k/n)} \operatorname{diag}\left(\mathbf{K}_\tau^{(n)}\right)' \mathbf{A}\left(\mathbf{s}_{\lceil t \rceil - 1}\right)\boldsymbol{\theta} \\ - \frac{1}{2} \sum_{\tau \in I(k/n)} \boldsymbol{\theta} A'\left(\mathbf{s}_{\lceil t \rceil - 1}\right)\mathbf{K}_{k/n,j,j}^{(n)}\mathbf{A}\left(\mathbf{s}_{\lceil t \rceil - 1}\right)\boldsymbol{\theta} \end{aligned}$$

is a discrete version of the quadratic in expression (4.5) of continuously traded wealth $W_t(\boldsymbol{\theta})$. (Recall that $I(k/n) \equiv \{1, \ldots, \lceil k/n \rceil - 1, k/n\}$.) Therefore, trading $\widehat{\mathbf{b}}_{k/n}$ achieves almost the same wealth as trading

$$\widehat{\mathbf{b}}^{(n)}_{k/n} = \frac{\int_{\mathbf{R}^d} \mathbf{b}\left(\boldsymbol{\theta}, \mathbf{s}_{\lceil k/n \rceil - 1}\right) \exp\left\{\mathbf{q}_{(k-1)/n}(\theta)\right\} \, d\pi(\boldsymbol{\theta})}{\int_{\mathbf{R}^d} \exp\left\{\mathbf{q}_{(k-1)/n}(\boldsymbol{\theta})\right\} \, d\pi(\boldsymbol{\theta})}. \tag{5.4}$$

However, using the form of $\mathbf{q}_{k/n}(\boldsymbol{\theta})$ we can quickly verify that measure

$$dG_{k,n}(\boldsymbol{\theta}) \equiv \frac{\exp\{\mathbf{q}_{k/n}(\boldsymbol{\theta})\} \, d\pi(\boldsymbol{\theta})}{\int_{\mathbf{R}^d} \exp\{\mathbf{q}_{k/n}(\boldsymbol{\theta})\} \, d\pi(\boldsymbol{\theta})}, \qquad \pi \sim N(\boldsymbol{\lambda}, \boldsymbol{\Lambda}), \tag{5.5}$$

is Gaussian with mean $\psi^{(n)}_{k/n}$ and covariance $\left(\boldsymbol{\Omega}^{(n)}_{k/n} + \boldsymbol{\Lambda}^{-1}\right)^{-1}$, and upon evaluating (5.4) we end up with the originally defined $\widehat{\mathbf{b}}^{(n)}_{k/n}$ of (5.2).

Key to the analysis is that $W_{(k-1)/n}(\boldsymbol{\theta})$ is closely approximated by the exponential quadratic $W_0 \exp\{\mathbf{q}_{(k-1)/n}(\boldsymbol{\theta})\}$ This can be concluded from the following lemma.

LEMMA 5.1. *Assume the minimal path condition holds. If at times $t \in \{k/n : k \in \{0, \ldots, Tn\}\}$ we rebalance according to $\mathbf{b}(\boldsymbol{\theta}, \mathbf{s}_{\lceil t \rceil - 1})$ with stock portfolio $\widetilde{\mathbf{b}}(\boldsymbol{\theta}, \mathbf{s}_{\lceil t \rceil - 1}) = \mathbf{A}(\mathbf{s}_{\lceil t \rceil - 1})\boldsymbol{\theta}$, then for constant $0 < C < 1$ and B such that $B < C \frac{\sqrt{n}}{L_p(1+\log n)}$ and $\boldsymbol{\theta}$ satisfying $\|\mathbf{A}(\mathbf{s}_{\lceil t \rceil - 1})\boldsymbol{\theta}\|_1 \le B$ for all t, the wealth achieved by time k/n is*

$$W^{(n)}_{k/n}(\boldsymbol{\theta}) = W_0 \exp\left\{\mathbf{q}_{k/n}(\boldsymbol{\theta}) + \varepsilon_{k/n}(\boldsymbol{\theta})\right\},$$

where $\varepsilon_{k/n}(\boldsymbol{\theta})$ is $O(B^3 k n^{-3/2} \log^3 n)$. Moreover, the distance between consecutive remainder terms, $|\varepsilon_{k/n}(\boldsymbol{\theta}) - \varepsilon_{(k-1)/n}(\boldsymbol{\theta})|$ is $O(B^3 n^{-3/2} \log^3 n)$.

Proof. See the Appendix.

The lemma is essentially a refinement of Theorem 4.1, which gives expressions for continuous time wealth $W_t(\boldsymbol{\theta})$. If $n \to \infty$ in the above lemma we can verify that quadratic $\mathbf{q}_{k/n}(\boldsymbol{\theta})$ converges to the same quadratics used in the expressions of $W_t(\boldsymbol{\theta})$ found in Theorem 4.1.

We now present a theorem showing that $\widehat{\mathbf{b}}^{(n)}_{k/n}$ achieves a growth rate of wealth arbitrarily close to that of the hindsight optimal wealth W^*_T of the continuously traded target class $(\mathbf{R}^d, S, \mathbf{A}(\mathbf{s}))$. This is predicated on the following conditions.

5.1.1. Universality Conditions (Discretized Case). For all $n \in \mathbf{N}$, $T \in \mathbf{N}$ and $k \in \{1, \ldots, Tn\}$, the following hold.

1. $\left(1 + \frac{L_p(1+\log n)}{\sqrt{n}}\right)^{-1} \le \frac{p_{k/n,j}}{p_{(k-1)/n,j}} \le 1 + \frac{L_p(1+\log n)}{\sqrt{n}}$ for some $L_p > 0$.
2. There exists constant $L_{\boldsymbol{\mu}} > 0$, independent of n and k such that, $\frac{1}{\lceil k/n \rceil} \sum_{\tau \in I(k/n)} \|\boldsymbol{\mu}^{(n)}_\tau\| < L_{\boldsymbol{\mu}}$. (Here, $I(k/n) \equiv \{1, \ldots, \lceil k/n \rceil - 1, k/n\}$.)
3. The empirical covariance matrix $\mathbf{K}^{(n)}_{k/n}$ is positive definite and there exists constant $L_{\mathbf{K}} > 0$ independent of n and k such that $\frac{1}{\lceil k/n \rceil} \sum_{\tau \in I(k/n)} \lambda_{\max}(\mathbf{K}^{(n)}_{k/n}) \le L_{\mathbf{K}}$.
4. There exists a constant $\widetilde{L}_{\mathbf{K}} > 0$ such that $|\lambda|_{\max}(\mathbf{K}_{k/n} - \mathbf{K}^{(n)}_{k/n}) \le \frac{\widetilde{L}_{\mathbf{K}}}{\sqrt{n}}$.
5. For any $\mathbf{s} \in S$, the $m \times d$ matrix $\mathbf{A}(\mathbf{s})$ is of full rank and there exist positive constants $L_{\mathbf{A},m}$ and $L_{\mathbf{A},d}$ such that $\lambda_{\max}(\mathbf{A}(\mathbf{s})\mathbf{A}'(\mathbf{s})) \le L_{\mathbf{A},m}$ and $\lambda_{\max}(\mathbf{A}'(\mathbf{s})\mathbf{A}(\mathbf{s})) \le L_{\mathbf{A},d}$.

6. The number of periods required for $\boldsymbol{\Omega}_{k/n}^{(n)}$ to become invertible is at most some integer β.
7. For all $k/n > \beta$ there exists positive constants L_{Ω}^{-} and L_{Ω}^{+} independent of T such that $L_{\Omega}^{-}(k/n - \beta) \leq \lambda_{\min}(\boldsymbol{\Omega}_{k/n}^{(n)})$ and $\lambda_{\max}(\boldsymbol{\Omega}_{k/n}^{(n)}) \leq L_{\Omega}^{+}(k/n - \beta)$.

THEOREM 5.1. *Assume that the above universality conditions (discretized case) hold. Let* $\widehat{W}_T^{(n)}$ *be the wealth achieved by trading portfolio sequence* $\widehat{\mathbf{b}}_{k/n}^{(n)}$ *of formula (5.2) to time T. Let* W_T^* *be the wealth achieved by the best in hindsight strategy in the continuously traded target class* $(\mathbf{R}^d, S, \mathbf{A}(\mathbf{s})\boldsymbol{\theta})$. *Then for any* $\epsilon > 0$ *we can find an integer N and positive constants* α *and C such that*

$$\frac{W_T^*}{\widehat{W}_T^{(n)}} \leq CT^{d/2} \exp\left\{\alpha T/n^{1/2-\varepsilon}\right\}$$

for all $n > N$.

Proof. See the Appendix.

Theorem 5.1 explains how we can use $\widehat{\mathbf{b}}_{k/n}^{(n)}$ to be universal with respect to the target wealth W_T^*. The first observation we make is that, for fixed n,

$$\begin{aligned} \lim_{T\to\infty} \sup_{p_t} \frac{1}{T} \log \frac{W_T^*}{\widehat{W}_T^{(n(T))}} &\leq \lim_{T\to\infty} \frac{1}{T} \log\left(CT^{d/2} \exp\left\{\alpha T/n^{1/2-\varepsilon}\right\}\right) \\ &= \alpha/n^{1/2-\varepsilon}. \end{aligned}$$

(Here, $\sup_{p_t}$ represents the supremum over all price paths satisfying the minimal path and universality conditions.) Hence $\widehat{\mathbf{b}}_{k/n}^{(n)}$ comes within a factor of $\alpha/n^{1/2-\varepsilon}$ of the optimal growth rate. However we can make this difference arbitrarily small by choosing n sufficiently large. To be actually universal though, we need to trade $\widehat{\mathbf{b}}_{k/n}^{(n)}$ at increasingly smaller intervals. By making $n(T)$ an increasing function of T such that $\lim_{T\to\infty} n(T) = \infty$, we see that

$$\lim_{T\to\infty} \sup_{p_t} \frac{1}{T} \log \frac{W_T^*}{\widehat{W}_T^{(n(T))}} \leq \lim_{T\to\infty} \alpha/n(T)^{1/2-\varepsilon} = 0,$$

which implies that $\widehat{\mathbf{b}}_{k/n}^{(n)}$ is now universal. The second observation we make is that if $n(T) \geq T^{2+\delta}$ for some $\delta > 0$ then the wealth ratio $W_T^*/\widehat{W}_T^{(n)}$ becomes $O(T^{d/2})$. To justify this, note that

$$\begin{aligned} \frac{W_T^*}{\widehat{W}_T^{(n(T))}} &\leq CT^{d/2} \exp\{\alpha T/n^{1/2-\varepsilon}\} \\ &\leq CT^{d/2} \exp\{\alpha T/T^{1-2\varepsilon+\delta/2-\varepsilon\delta}\}. \end{aligned}$$

By choosing $\varepsilon < \frac{\delta}{2(2+\delta)}$, it follows that $1 - 2\varepsilon + \delta/2 - \varepsilon\delta > 1$ and hence $T/T^{1-2\varepsilon+\delta/2-\varepsilon\delta}$ is upper bounded by some constant, implying that

$$\frac{W_T^*}{\widehat{W}_T^{(n(T))}} \leq CT^{d/2}.$$

Hence $\widehat{\mathbf{b}}_{k/n}^{(n)}$ comes within a polynomial bound of the target wealth on this schedule.

5.2. Final Note on Computation

A nice property of $\widehat{\mathbf{b}}^{(n)}_{k/n}$ is that it is computable in order $\max\{m^2, d^2\}$ steps where m is the number of stocks and d is the dimension of the target class parameter space. To see this note that computation at each step depends on computing $\psi^{(n)}_{(k-1)/n}$. To compute $\psi^{(n)}_{(k-1)/n}$ we start with stored values of $\boldsymbol{\Omega}^{(n)}_{(k-2)/n}$ and $\mathbf{K}^{(n)}_{(k-2)/n}$ and note that

$$\mathbf{K}^{(n)}_{(k-1)/n} = \mathbf{K}^{(n)}_{(k-2)/n} + \boldsymbol{\delta}_{(k-1)/n}\boldsymbol{\delta}'_{(k-1)/n} \tag{5.7}$$

and

$$\boldsymbol{\Omega}^{(n)}_{(k-1)/n} = \boldsymbol{\Omega}^{(n)}_{(k-2)/n} + \mathbf{A}'\left(\mathbf{s}_{\lceil k/n \rceil - 1}\right)\boldsymbol{\delta}_{(k-1)/n}\boldsymbol{\delta}'_{(k-1)/n}\mathbf{A}\left(\mathbf{s}_{\lceil k/n \rceil - 1}\right), \tag{5.8}$$

where $\boldsymbol{\delta}_{(k-1)/n}$ is the m-dimensional incremental log drift vector

$$\boldsymbol{\delta}_{(k-1)/n} = Z_{(k-1)/n} - Z_{(k-2)/n}.$$

Both updates (5.7) and (5.8) are computable on order $\max\{m^2, d^2\}$. The next quantity needed to compute $\psi^{(n)}_{(k-1)/n}$ is the inverse $\left(\boldsymbol{\Omega}^{(n)}_{(k-1)/n} + \boldsymbol{\Lambda}^{-1}\right)^{-1}$. Generally the computation of the inverse of a $d \times d$ matrix such as $\boldsymbol{\Omega}^{(n)}_{(k-1)/n} + \boldsymbol{\Lambda}^{-1}$ requires on the order of d^3 steps. However if we store $\left(\boldsymbol{\Omega}^{(n)}_{(k-2)/n} + \boldsymbol{\Lambda}^{-1}\right)^{-1}$ it is possible to use the Sherman-Morrison formula for matrix inversion (see Golub and Van Loan, 1997) to reduce the number of steps to order d^2. In particular, if M is any invertible $d \times d$ matrix and if $\mathbf{v}$ is any d dimensional vector then it is true that

$$(M + \mathbf{v}\mathbf{v}')^{-1} = M^{-1} + \frac{M^{-1}\mathbf{v}\mathbf{v}'M^{-1}}{1 + \mathbf{v}'M^{-1}\mathbf{v}}.$$

Since the right-hand side computes on order d^2 steps we can set M to $\boldsymbol{\Omega}^{(n)}_{(k-2)/n} + \boldsymbol{\Lambda}^{-1}$ and $\mathbf{v}$ to $\boldsymbol{\delta}_{(k-1)/n}$ to conclude that $\left(\boldsymbol{\Omega}^{(n)}_{(k-2)/n} + \boldsymbol{\Lambda}^{-1}\right)^{-1}$ is computable on order d^2 steps. Thus every component of $\psi^{(n)}_{(k-1)/n}$ is computable at the most on order $\max\{m^2, d^2\}$ steps, so it follows that $\widehat{\mathbf{b}}^{(n)}_{k/n}$ is as well.

6. CONCLUSION

In this paper we have developed a discretely traded universal portfolio (i.e., $\widehat{\mathbf{b}}^{(n)}_{k/n}$ of (5.2)) that, under certain assumptions, achieves the same growth rate as the best hindsight strategy among a user-specified, continuously traded target class. The portfolio offers two main advantages over other previous procedures. The first advantage is that the algorithm used to generate the universal portfolio can be adapted so that the resulting portfolio can be universal with respect to different linearly parameterized target classes. The target classes may also employ a continuous form of dependence on side information not previously considered. In this setting the continuous rebalanced portfolio class used by previous authors becomes a special case.

The second advantage of our portfolio is that, unlike previous procedures, the portfolio is computable in a polynomial time through a simple closed-form expression. The motivation of the expression borrows heavily from the Bayesian concept of normal conjugation. Knowing that the wealth function of continuously traded linearly parameterized target classes is Gaussian (à la (4.3)) and that the universal portfolio is merely a wealth-weighted combination of the target class portfolios, if one averages these portfolios using a second

Gaussian measure, the resulting universal portfolio is interpretable as the mean of a "posterior" Gaussian distribution for which elementary closed-form expressions exist. This is an improvement over other procedures, which in general require numerical methods to compute.

As a final note, we reiterate that obtaining an efficient universal procedure for more general target classes comes at a cost in that universality is restricted to sets of price paths satisfying certain regularity conditions, the most notable being that prices are bounded above zero. When restricting consideration to the constant rebalanced portfolio target class considered by Cover (1991) such regularity conditions are generally unnecessary to achieve a universal procedure. Indeed this also true for our discrete time results in the third section of the paper if we choose to use constant rebalanced portfolios as our target class. This follows from noting that under this choice of target class our framework reduces to a direct analog of the framework originally proposed by Cover. However for the more general target classes considered in the bulk of the paper, the scope of universality is reduced to a more restrictive set of price paths satisfying the given regularity conditions.

APPENDIX: PROOFS

A.1 Proof of Theorem 3.1

To prove Theorem 3.1 we use the following lemma.

LEMMA A.1. *Under the assumptions of Theorem 3.1, there exists a constant $R > 0$ independent of n such that $\pi(\Theta_n^*) \geq Rn^{-d}$.*

Proof. First note that because the density of π is uniformly bounded above 0 by $\delta > 0$ it follows that $\pi(\Theta_n^*) \geq \delta \operatorname{Vol}(\Theta_n^*)$, where $\operatorname{Vol}(\cdot)$ denotes the Lebesgue measure (or volume) of a set.

To bound $\operatorname{Vol}(\Theta_n^*)$ it is useful to use the identity $\Theta_n^* = \Theta \bigcap B(\boldsymbol{\theta}_n^*, \frac{1}{n})$, where $B(\boldsymbol{\theta}_n^*, \frac{1}{n}) = \{\boldsymbol{\theta} \in \mathbf{R}^d : \|\boldsymbol{\theta} - \boldsymbol{\theta}_n^*\| \leq \frac{1}{n}\}$ is a $\boldsymbol{\theta}_n^*$ centered closed ball in $\mathbf{R}^d$. Using this identity we endeavor to show that there exists a constant $C > 0$ independent of n for which $\operatorname{Vol}(\Theta_n^*) \geq C \operatorname{Vol}(B(\boldsymbol{\theta}_n^*, \frac{\epsilon}{n}))$. This can be shown by taking a d-dimensional closed ball in Θ and considering its convex extension to $\boldsymbol{\theta}_n^*$. By examining the volume of the intersection between this convex hull and $B(\boldsymbol{\theta}_n^*, \frac{1}{n})$ we are able to show that C exists.

First, to justify the existence of a d-dimensional closed ball in Θ we note that there exist d points in Θ such that there is no $d-1$ dimensional hyperplane containing all d points. If this were not the case, Θ would lie in a $d-1$ subspace and thus would have a null Lebesgue measure, which contradicts our assumptions. By the convexity of Θ, the convex hull of these d points is a subset of Θ and clearly a closed ball exists in such a hull.

Suppose this closed ball has center $\boldsymbol{\theta}_0$ and radius $r > 0$. Label it $B(\boldsymbol{\theta}_0, r) \equiv \{\boldsymbol{\theta} \in R^d \,|\, \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq r\}$. Now define a convex extension to this ball. For parameter $\boldsymbol{\theta}_n^* \in \Theta$ we define $H(\boldsymbol{\theta}_n^*) = \{\boldsymbol{\theta} : \boldsymbol{\theta} = \lambda \boldsymbol{\theta}_n^* + (1 - \lambda)\boldsymbol{\theta}', \lambda \in [0, 1], \boldsymbol{\theta}' \in B(\boldsymbol{\theta}_0, r)\}$. By its definition, $H(\boldsymbol{\theta}_n^*)$ is the convex hull of $B(\boldsymbol{\theta}_0, r)$ and maximal parameter $\boldsymbol{\theta}_n^*$. The set can be visualized as an ice cream cone with tip $\boldsymbol{\theta}_n^*$ and scoop $B(\boldsymbol{\theta}_0, r)$. Since $\boldsymbol{\theta}^*$ is contained in Θ as is $B(\boldsymbol{\theta}_0, r)$, it follows from convexity of Θ that $H(\boldsymbol{\theta}_n^*)$ is also in Θ.

Now consider the volume of the intersection between ball $B(\boldsymbol{\theta}_n^*, 1/n)$ and cone $H(\boldsymbol{\theta}_n^*)$. The cone $H(\boldsymbol{\theta}_n^*)$ has been purposely defined to have its tip coincide with the center of

$B(\boldsymbol{\theta}_n^*, 1/n)$. For sufficiently large n, the radius of $B(\boldsymbol{\theta}_n^*, 1/n)$ will be smaller than that of $B(\boldsymbol{\theta}_0, r)$, the ball atop $H(\boldsymbol{\theta}_n^*)$. In this case, a geometric argument shows that

$$\mathrm{Vol}\left(B(\boldsymbol{\theta}_n^*, 1/n) \bigcap H(\boldsymbol{\theta}_n^*)\right) \geq \mathbf{A}\left(\boldsymbol{\theta}_n^*\right) \mathrm{Vol}\left(B\left(\boldsymbol{\theta}_n^*, 1/n\right)\right),$$

where $\mathbf{A}(\boldsymbol{\theta}_n^*)$ is the proportion of the surface area of $B(\boldsymbol{\theta}_n^*, 1/n)$ contained in $H(\boldsymbol{\theta}_n^*)$. As $\boldsymbol{\theta}_n^*$ gets farther away from $\boldsymbol{\theta}_0$ (the center of the scoop), the cone narrows and $\mathbf{A}(\boldsymbol{\theta}_n^*)$ gets smaller. However it only vanishes completely when this distance between $\boldsymbol{\theta}_n^*$ and $\boldsymbol{\theta}_0$ is infinite. Since Θ is compact, the distance is bounded and hence $\mathbf{A}(\boldsymbol{\theta}_n^*)$ is uniformly bounded above 0 for all $\boldsymbol{\theta}_n^* \in \Theta$. Thus we can select constant $C > 0$ such that

$$\text{(A.1)} \qquad \mathrm{Vol}\left(B\left(\boldsymbol{\theta}_n^*, 1/n\right) \bigcap H\left(\boldsymbol{\theta}_n^*\right)\right) \geq C\,\mathrm{Vol}\left(B\left(\boldsymbol{\theta}_n^*, 1/n\right)\right), \qquad \forall \boldsymbol{\theta}_n^* \in \Theta^*.$$

Although this inequality is only justified for n larger than some $N > 0$, we can make (A.1) hold for all n by defining C to be the lesser of

$$\inf_{n \leq N} \left\{ \frac{\mathrm{Vol}\left(B\left(\boldsymbol{\theta}_n^*, 1/n\right) \bigcap H\left(\boldsymbol{\theta}_n^*\right)\right)}{\mathrm{Vol}\left(H\left(\boldsymbol{\theta}^*\right)\right)} \right\},$$

and

$$\inf_{\boldsymbol{\theta}_n^* \in \Theta, n > N} \mathbf{A}\left(\boldsymbol{\theta}_n^*\right).$$

Both these infimums are strictly positive so we have $C > 0$ as required.

To end the proof we note that

$$\begin{aligned}
\pi\left(\boldsymbol{\theta}_{\epsilon/n}^*\right) &\geq \delta\,\mathrm{Vol}\left(\Theta_n^*\right) \\
&= \delta\,\mathrm{Vol}\left(B\left(\boldsymbol{\theta}_n^*, 1/n\right) \bigcap \Theta\right) \\
&\geq \delta\,\mathrm{Vol}\left(B\left(\boldsymbol{\theta}_n^*, 1/n\right) \bigcap H\left(\boldsymbol{\theta}_n^*\right)\right) \\
&\geq \delta\, C \mathrm{Vol}\left(B\left(\boldsymbol{\theta}_n^*, 1/n\right)\right) \\
&= \delta C \mathbf{K} n^{-d} \qquad \text{for some } \mathbf{K} > 0 \\
&= R n^{-d} \qquad \text{with } R = \delta C \mathbf{K}.
\end{aligned}$$

□

A.2 Proof of Theorem 4.1 and Lemma 5.1

Theorem 4.1 follows from Lemma 5.1 so we prove the latter first.

Proof. Suppose we trade $\mathbf{b}(\boldsymbol{\theta}, \mathbf{s})$ a total of n times a period for T periods. For wealth relatives $\mathbf{x}_{h/n} = (x_{h/n,0}, x_{h/n,1}, \ldots, x_{h/n,m})$, with $\mathbf{x}_{h/n,j} = p_{h/n,j}/p_{(h-1)/n,j}$ denoting the wealth relative for the jth stock and $x_{h/n,0} = 1$ denoting the wealth relative of cash, the wealth achieved by this strategy by time $t = k/n$ is

$$W_{k/n}^{(n)}(\boldsymbol{\theta}) = W_0 \prod_{h=1}^{k} \mathbf{b}'\big(\boldsymbol{\theta}, \mathbf{s}_{\lceil h/n \rceil - 1}\big) \mathbf{x}_{h/n}.$$

For simplification we hereafter write $\mathbf{b}_{h/n}$ for $\mathbf{b}'(\boldsymbol{\theta}, \mathbf{s}_{\lceil h/n \rceil - 1})$ with the understanding that $\mathbf{b}_{h/n}$ is dependent on $\boldsymbol{\theta}$ and side information state $\mathbf{s}_{\lceil h/n \rceil - 1}$. Now we define $r_{h/n,j} \equiv$

$\log(x_{h/n,j})$ and write

$$W_{k/n}^{(n)}(\boldsymbol{\theta}) = W_0 \prod_{h=1}^{k} \sum_{j=0}^{m} b_{h/n,j} \exp\{r_{h/n,j}\}$$
$$= W_0 \exp\left\{ \sum_{h=1}^{k} \log\left(\sum_{j=0}^{m} b_{h/n,j} \exp\{r_{h/n,j}\} \right) \right\}.$$

We interpret the exponent of (A.2) as a function of $(r_{h/n,1}, \dots, r_{h/n,m})$, (noting that $r_{h/n,0}$ is always *0* because $x_{h/n,0} = 1$), and we define $f(r_{h/n,1}, \dots, r_{h/n,m}) = \log(\sum_{j=0}^{m} b_{h/n,j} \exp\{r_{h/n,j}\})$. Taking the Taylor expansion of f about $(0, \dots, 0)$ we get

(A.3)

$$f(r_{h/n,1}, \dots, r_{h/n,m}) = \sum_{j=1}^{m} b_{h/n,j} r_{h/n,j} + \frac{1}{2} \sum_{j=1}^{m} b_{h/n,j} r_{h/n,j}^2 - \frac{1}{2} \left(\sum_{j=1}^{m} b_{h/n,j} r_{h/n,j} \right)^2$$
$$+ \frac{1}{6} \sum_{i,j,k=1}^{m} \left(\frac{\partial^3 f}{\partial r_{h/n,i} \partial r_{h/n,j} \partial r_{h/n,k}} (c_1, \dots, c_m) \right) r_{h/n,i} r_{h/n,j} r_{h/n,k}$$

for some $(c_1, \dots, c_m)$ between $(r_{h/n,1}, \dots, r_{h/n,m})$ and $(0, \dots, 0)$. In order to bound the remainder term in the expansion we first claim that

$$\text{(A.4)} \quad \sum_{i,j,k=1}^{m} \frac{\partial^3 f}{\partial r_{h/n,i} \partial r_{h/n,j} \partial r_{h/n,k}} (c_1, \dots, c_m) = 2 \sum_{i,j,k=1}^{m} \frac{b_{h/n,i} e^{c_i} b_{h/n,j} e^{c_j} b_{h/n,k} e^{c_k}}{\left(b_0 + \sum_{j=1}^{m} b_{h/n,j} e^{c_j} \right)^3}$$
$$- 3 \sum_{i,j=1}^{m} \frac{b_{h/n,i} e^{c_i} b_{h/n,j} e^{c_j}}{\left(b_0 + \sum_{j=1}^{m} b_{h/n,j} e^{c_j} \right)^2}$$
$$+ \sum_{i,j=1}^{m} \frac{b_{h/n,i} e^{c_i}}{b_0 + \sum_{j=1}^{m} b_{h/n,j} e^{c_j}}.$$

But note that since $\|\mathbf{A}(\mathbf{s}_{\lfloor h/n \rfloor + 1})\boldsymbol{\theta}\|_1 = \sum_{j=1}^{m} |b_{h/n,j}| \le B$ it follows that

$$\text{(A.5)} \quad \left| b_{h/n,i} e^{c_i} \right| \le B \max\left\{ p_{h/n,j} / p_{(h-1)/n,j}, 1 \right\}$$
$$\le B \left(1 + \frac{L_p (1 + \log n)}{\sqrt{n}} \right) = O(B).$$

Also note that

$$b_0 + \sum_{j=1}^{m} b_{h/n,j} e^{c_j} = 1 + \sum_{j=1}^{m} b_{h/n,j} \left(\frac{p_{h/n,j}}{p_{(h-1)/n,j}} - 1 \right)$$
$$\ge 1 - B \left(\frac{L_p (1 + \log n)}{\sqrt{n}} \right).$$

Since B is defined to be strictly less than $C \frac{\sqrt{n}}{L_p(1+\log n)}$ for constant C less than 1 it follows from the above inequality that there exists some positive constant C' independent of n

such that

$$b_0 + \sum_{j=1}^{m} b_{h/n,j} e^{c_j} \geq C'. \tag{A.6}$$

Using (A.5) and (A.6) we return to (A.4) and conclude that

$$\sum_{i,j,k=1}^{m} \frac{\partial^3 f}{\partial r_{h/n,i} \partial r_{h/n,j} \partial r_{h/n,k}} (c_1, \ldots, c_m) = O(B^3).$$

Moreover, since $|r_{h/n,j}| = |\log(p_{h/n,j}/p_{(h-1)/n,j})| \leq \frac{L_p(1+\log n)}{\sqrt{n}} = O(n^{-1/2} \log n)$ it follows from this and the above equation that the remainder term in (A.3) is $O(B^3 n^{-3/2} \log^3 n)$. Hence we can write

$$\begin{aligned} f(r_{h/n,1}, \ldots, r_{h/n,m}) = & \sum_{j=1}^{m} b_{h/n,j} r_{h/n,j} + \frac{1}{2} \sum_{j=1}^{m} b_{h/n,j} r_{h/n,j}^2 \\ & - \frac{1}{2} \left(\sum_{j=1}^{m} b_{h/n,j} r_{h/n,j} \right)^2 + O(B^3 n^{-3/2} \log^3 n). \end{aligned} \tag{A.7}$$

Recalling that we originally set $f(r_{h/n,1}, \ldots, r_{h/n,m}) = \log\left(\sum_{j=0}^{m} b_{h/n,j} \exp\{r_{h/n,j}\}\right)$ we return to (A.2) and use (A.7) to conclude that

$$\begin{aligned} W_{k/n}^{(n)}(\boldsymbol{\theta}) = W_0 \exp \Bigg\{ & \sum_{h=1}^{k} \sum_{j=1}^{m} b_{h/n,j} r_{h/n,j} + \frac{1}{2} \sum_{h=1}^{k} \sum_{j=1}^{m} b_{h/n,j} r_{h/n,j}^2 \\ & - \frac{1}{2} \sum_{h=1}^{k} \left(\sum_{j=1}^{m} b_{h/n,j} r_{h/n,j} \right)^2 + O(B^3 k n^{-3/2} \log^3 n) \Bigg\}. \end{aligned} \tag{A.8}$$

Now note that

$$\begin{aligned} \sum_{h=1}^{k} \sum_{j=1}^{m} b_{h/n,j} r_{h/n,j} &= \sum_{\tau \in I(k/n)} \sum_{j=1}^{m} b_j\left(\boldsymbol{\theta}, \mathbf{s}_{\lceil \tau \rceil - 1}\right) \left(Z_{\tau,j} - Z_{\lceil \tau \rceil - 1, j}\right) \\ &= \sum_{\tau \in I(k/n)} \boldsymbol{\mu}_t' \widetilde{b}\left(\boldsymbol{\theta}, \mathbf{s}_{\lceil \tau \rceil - 1}\right) \\ &= \sum_{\tau \in I(k/n)} \boldsymbol{\mu}_t' \mathbf{A}\left(\mathbf{s}_{\lceil \tau \rceil - 1}\right) \boldsymbol{\theta} \end{aligned} \tag{A.9}$$

and

$$\begin{aligned} \frac{1}{2} \sum_{h=1}^{k} \sum_{j=1}^{m} b_{h/n,j} r_{h/n,j}^2 &= \frac{1}{2} \sum_{h=1}^{k} \sum_{j=1}^{m} b_j\left(\boldsymbol{\theta}, \mathbf{s}_{\lceil h/n \rceil - 1}\right) \left(Z_{(h/n),j} - Z_{((h-1)/n),j}\right)^2 \\ &= \frac{1}{2} \sum_{\tau \in I(k/n)} \sum_{j=1}^{m} b_j\left(\boldsymbol{\theta}, \mathbf{s}_{\lceil \tau \rceil - 1}\right) \mathbf{K}_{\tau,j,j}^{(n)} \\ &= \frac{1}{2} \sum_{\tau \in I(k/n)} \operatorname{diag}\left(\mathbf{K}_{\tau}^{(n)}\right)' \widetilde{b}\left(\boldsymbol{\theta}, \mathbf{s}_{\lceil \tau \rceil - 1}\right) \\ &= \frac{1}{2} \sum_{\tau \in I(k/n)} \operatorname{diag}\left(\mathbf{K}_{\tau}^{(n)}\right)' \mathbf{A}\left(\mathbf{s}_{\lceil \tau \rceil - 1}\right) \boldsymbol{\theta} \end{aligned} \tag{A.10}$$

and, finally, that

$$\text{(A.11)}\quad \frac{1}{2}\sum_{h=1}^{k}\left(\sum_{j=1}^{m} b_{h/n,j} r_{h/n,j}\right)^2 = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} b_{h/n,i} b_{h/n,j} \sum_{k=1}^{n}\left(Z_{(k/n),i} - Z_{((k-1)/n),i}\right) \times \left(Z_{(k/n),j} - Z_{((k-1)/n),j}\right)$$
$$= \frac{1}{2}\sum_{\tau\in I(k/n)} \widetilde{b}'(\boldsymbol{\theta}, \mathbf{s}_{\lceil\tau\rceil-1})\mathbf{K}_\tau^{(n)}\widetilde{b}(\boldsymbol{\theta}, \mathbf{s}_{\lceil\tau\rceil-1})$$
$$= \frac{1}{2}\sum_{\tau\in I(k/n)} \boldsymbol{\theta} A'\left(\mathbf{s}_{\lceil\tau\rceil-1}\right)\mathbf{K}_{k/n,j,j}^{(n)}\mathbf{A}\left(\mathbf{s}_{\lceil\tau\rceil-1}\right)\boldsymbol{\theta}.$$

Lemma 5.1 is proved upon substituting (A.9), (A.10), and (A.11) into (A.8) and setting $\varepsilon_{k/n}(\boldsymbol{\theta})$ to be the $O(B^3 k n^{-3/2}\log^3 n)$ term of (7.8).

As for Theorem 4.1, expression (4.3) is obtained by taking the results of Lemma 5.1 and letting $n\to\infty$. The second expression for $W_T(\boldsymbol{\theta})$, (4.4), is obtained upon completing the square in (4.3). □

6.1. Proof of Theorem 5.1

Proof. At various points in the proof we will work with tails of Gaussian distributions. For this reason we will say that a function $f(x)$ is $O_{gt}(x)$ (or of Gaussian tail order) if there exist positive constants C_1 and C_2 such that $f(x)\le C_1\exp\{-C_2x^2\}$. Recall the Gaussian measure

$$\text{(A.12)}\qquad dG_{k,n}(\boldsymbol{\theta}) \equiv \frac{\exp\left\{\mathbf{q}_{k/n}(\boldsymbol{\theta})\right\}\, d\pi(\boldsymbol{\theta})}{\int_{\mathbf{R}^d}\exp\left\{\mathbf{q}_{k/n}(\boldsymbol{\theta})\right\}\, d\pi(\boldsymbol{\theta})},\qquad \pi\sim N(\boldsymbol{\lambda},\boldsymbol{\Lambda}).$$

By completing the square in the exponent of the density of $G_{k,n}$ it can be verified that $G_{k,n}$ has mean

$$\psi_{k/n}^{(n)} \equiv \left(\boldsymbol{\Omega}_{k/n}^{(n)} + \boldsymbol{\Lambda}^{-1}\right)^{-1} \times\left(\mathbf{A}'(\mathbf{s}_0)\left[\boldsymbol{\mu}_1 + \frac{1}{2}\text{diag}\left(\mathbf{K}_1^{(n)}\right)\right] + \cdots + \mathbf{A}'\left(\mathbf{s}_{\lceil k/n\rceil-1}\right)\left[\boldsymbol{\mu}_{k/n} + \frac{1}{2}\text{diag}\left(\mathbf{K}_{k/n}^{(n)}\right)\right]\right) + \boldsymbol{\Lambda}^{-1}\boldsymbol{\lambda}$$

and variance $\left(\boldsymbol{\Omega}_{k/n}^{(n)} + \boldsymbol{\Lambda}^{-1}\right)^{-1}$.

We argue that $\left\|\psi_{k/n}^{(n)}\right\|$ is bounded by some constant L_ψ independent of n and k. First consider times $k/n \le \beta+1$ where β is the maximum number of periods before $\boldsymbol{\Omega}_t^{(n)}$ becomes invertible. In this case we can use the universality conditions in showing

$$\text{(A.13)}\quad \left\|\psi_{k/n}^{(n)}\right\| \le \lambda_{\max}\left(\left(\boldsymbol{\Omega}_{k/n}^{(n)} + \boldsymbol{\Lambda}^{-1}\right)^{-1}\right)\sum_{\tau\in I(k/n)}\left\|\mathbf{A}'\left(\mathbf{s}_{\lceil\tau\rceil-1}\right)\left[\mu_\tau + \frac{1}{2}\,\text{diag}\left(\mathbf{K}_\tau^{(n)}\right)\right]\right\|$$
$$\le \frac{1}{\lambda_{\min}\left(\boldsymbol{\Lambda}^{-1}\right)} L_{\mathbf{A},m}^{1/2}\sum_{\tau\in I(k/n)}\left\|\boldsymbol{\mu}_\tau + \frac{1}{2}\,\text{diag}\left(\mathbf{K}_\tau^{(n)}\right)\right\|$$
$$\le \frac{L_{\mathbf{A},m}^{1/2}}{\lambda_{\min}\left(\boldsymbol{\Lambda}^{-1}\right)}\left(L_{\boldsymbol{\mu}} + \frac{\sqrt{m}}{2}L_{\mathbf{K}}\right)(\beta+1).$$

Similarly, if $k/n > \beta + 1$ then

$$\text{(A.14)} \quad \begin{aligned} \left\|\psi_{k/n}^{(n)}\right\| &\le \lambda_{\max}\left(\left(\mathbf{\Omega}_{k/n}^{(n)} + \mathbf{\Lambda}^{-1}\right)^{-1}\right) \sum_{\tau \in I(k/n)} \left\|\mathbf{A}'\left(\mathbf{s}_{\lceil\tau\rceil-1}\right)\left[\boldsymbol{\mu}_\tau + \frac{1}{2}\operatorname{diag}\left(\mathbf{K}_\tau^{(n)}\right)\right]\right\| \\ &\le \frac{1}{\lambda_{\min}\left(\mathbf{\Omega}_{k/n}^{(n)}\right)} L_{\mathbf{A},m}^{1/2} \sum_{\tau \in I(k/n)} \left\|\boldsymbol{\mu}_\tau + \frac{1}{2}\operatorname{diag}\left(\mathbf{K}_\tau^{(n)}\right)\right\| \\ &\le \frac{L_{\mathbf{A},m}^{1/2}}{L_{\mathbf{\Omega}}^{-}(k/n - \beta)}\left(L_\mu + \frac{\sqrt{m}}{2} L_{\mathbf{K}}\right)(k/n) \\ &\le \frac{L_{\mathbf{A},m}^{1/2}}{L_{\mathbf{\Omega}}^{-}}\left(L_\mu + \frac{\sqrt{m}}{2} L_{\mathbf{K}}\right)(\beta + 1). \end{aligned}$$

Therefore, for all n and k, $\|\psi_{k/n}^{(n)}\|$ is bounded by L_ψ where L_ψ is set to the greater of (A.13) and (A.14).

Now we define the set

$$\Theta_n \equiv \left\{\boldsymbol{\theta} \in \mathbf{R}^d : \|\boldsymbol{\theta}\| \le C\frac{n^{\epsilon/4}}{\sqrt{m} L_p L_{\mathbf{A},d}^{1/2}}\right\},$$

where for the moment C is the same constant as that defined in Lemma 5.1. The bound on $\|\boldsymbol{\theta}\|$ in the definition of Θ_n is purposely set so that Lemma 5.1 holds. To see this, note that, for each n and $\boldsymbol{\theta} \in \Theta_n$,

$$\begin{aligned} \left\|\mathbf{A}\left(\mathbf{s}_{\lceil k/n\rceil-1}\right)\boldsymbol{\theta}\right\|_1 &\le \sqrt{m}\left\|\mathbf{A}\left(\mathbf{s}_{\lceil k/n\rceil-1}\right)\boldsymbol{\theta}\right\| \\ &\le \sqrt{m} L_{\mathbf{A},d}^{1/2}\|\boldsymbol{\theta}\| \\ &\le \frac{n^{\epsilon/4}}{L_p} \\ &\le \frac{\sqrt{n}}{L_p(1 + \log n)}, \end{aligned}$$

and hence the L_1 condition of Lemma 5.1 is satisfied. Also, N as given in the statement is set sufficiently large to ensure that the bound on $\|\boldsymbol{\theta}\|$ is larger than L_ψ so that $\psi_{k/n}^{(n)} \in \Theta_n$ for all k and $n \ge N$.

At various points in the proof we will need a bound on $1 - \int_{\Theta_n} dG_{k,n}(\boldsymbol{\theta})$—that is, a bound on the tails of $G_{k,n}(\boldsymbol{\theta})$. Regardless of n and k the "fatness" of a tail is limited by the minimum eigenvalue of its covariance matrix $\left(\mathbf{\Omega}_{k/n}^{(n)} + \mathbf{\Lambda}^{-1}\right)^{-1}$ which in turn is bounded by the minimum eigenvalue of "prior" covariance $\mathbf{\Lambda}$. Thus the density of $G_{k,n}(\boldsymbol{\theta})$ cannot get arbitrarily flat with n or k. Also since $\|\psi_{k/n}^{(n)}\| < L_\psi$ the "peak" of $G_{k,n}(\boldsymbol{\theta})$ at mean $\psi_{k/n}^{(n)}$ is restricted to a distance of L_ψ from the origin. Because $G_{k,n}(\boldsymbol{\theta})$ cannot get arbitrarily flat and because its peak is always within a bounded distance from the origin we can find another flatter Gaussian distribution $Z(\boldsymbol{\theta})$ centered on the origin with constant covariance such that

$$1 - \int_{\Theta_n} dG_{k,n}(\boldsymbol{\theta}) \le 1 - \int_{\Theta_n} dZ(\boldsymbol{\theta})$$

for all n and k. But from the properties of Gaussian tails we know that $1 - \int_{\Theta_n} dZ(\boldsymbol{\theta})$ is $O_{gt}(n^{\epsilon/4})$ where $n^{\epsilon/4}$ is the order at which Θ_n expands. Hence,

$$1 - \int_{\Theta_n} dG_{k,n}(\boldsymbol{\theta}) = \int_{\Theta_n^C} dG_{k,n}(\boldsymbol{\theta}) = O_{gt}(n^{\epsilon/4}). \tag{A.15}$$

To continue, note that

$$\left\| \int_{\Theta_n^C} \boldsymbol{\theta}\, dG_{k,n}(\boldsymbol{\theta}) \right\| \le \int_{\Theta_n^C} \|\boldsymbol{\theta}\| \frac{dG_{k,n}(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\, d\boldsymbol{\theta}. \tag{A.16}$$

Since density $\frac{dG_{k,n}(\boldsymbol{\theta})}{d\boldsymbol{\theta}}$ is $O_{gt}(\|\boldsymbol{\theta}\|)$ it follows that $\|\boldsymbol{\theta}\| \frac{dG_{k,n}(\boldsymbol{\theta})}{d\boldsymbol{\theta}}$ is also $O_{gt}(\|\boldsymbol{\theta}\|)$. Upon noting that

$$\int_{\Theta_n^C} O_{gt}(\|\boldsymbol{\theta}\|) = O_{gt}(n^{\epsilon/4}),$$

we conclude that

$$\begin{aligned} \left\| \int_{\Theta_n^C} \boldsymbol{\theta}\, dG_{k,n}(\boldsymbol{\theta}) \right\| &\le \int_{\Theta_n^C} \|\boldsymbol{\theta}\| \frac{dG_{k,n}(\boldsymbol{\theta})}{d\boldsymbol{\theta}} d\boldsymbol{\theta} \\ &= \int_{\Theta_n^C} O_{gt}(\|\boldsymbol{\theta}\|) \\ &= O_{gt}(n^{\epsilon/4}). \end{aligned} \tag{A.17}$$

Now we use bounds (A.15) and (A.17) to compare $\psi_{k/n}^{(n)}$ to

$$\psi_{k/n}^{\dagger(n)} \equiv \frac{\int_{\Theta_n} \boldsymbol{\theta}\, dG_{k,n}(\boldsymbol{\theta})}{\int_{\Theta_n} dG_{k,n}(\boldsymbol{\theta})}.$$

and bound the distance $\|\psi_{k/n}^{(n)} - \psi_{k/n}^{\dagger(n)}\|$. Since $\psi_{k/n}^{(n)}$ is the mean of $G_{k,n}(\boldsymbol{\theta})$ we write $\psi_{k/n}^{(n)} = \int_{\mathbf{R}^d} \boldsymbol{\theta} dG_{k,n}(\boldsymbol{\theta})$ and hence,

$$\begin{aligned} \left\| \psi_{k/n}^{(n)} - \psi_{k/n}^{\dagger(n)} \right\| &= \left\| \int_{\mathbf{R}^d} \boldsymbol{\theta}\, dG_{k,n}(\boldsymbol{\theta}) - \frac{\int_{\Theta_n} \boldsymbol{\theta}\, dG_{k,n}(\boldsymbol{\theta})}{\int_{\Theta_n} dG_{k,n}(\boldsymbol{\theta})} \right\| \\ &= \left\| \frac{\int_{\Theta_n^C} \boldsymbol{\theta}\, dG_{k,n}(\boldsymbol{\theta})}{\int_{\Theta_n} dG_{k,n}(\boldsymbol{\theta})} - \frac{\int_{\Theta_n^C} dG_{k,n}(\boldsymbol{\theta})}{\int_{\Theta_n} dG_{k,n}(\boldsymbol{\theta})} \int_{\mathbf{R}^d} \boldsymbol{\theta}\, dG_{k,n}(\boldsymbol{\theta}) \right\| \\ &\le \left\| \frac{\int_{\Theta_n^C} \boldsymbol{\theta}\, dG_{k,n}(\boldsymbol{\theta})}{\int_{\Theta_n} dG_{k,n}(\boldsymbol{\theta})} \right\| + \left\| \frac{\int_{\Theta_n^C} dG_{k,n}(\boldsymbol{\theta})}{\int_{\Theta_n} dG_{k,n}(\boldsymbol{\theta})} \right\| L_\psi \\ &= O_{gt}(n^{\epsilon/4}) + O_{gt}(n^{\epsilon/4}) \\ &= O_{gt}(n^{\epsilon/4}) \end{aligned} \tag{A.18}$$

(by (A.15) and (A.17)).

Now let $\mathbf{x}_{k/n}$ denote the vector of wealth relatives at time k/n,

$$\begin{aligned}\mathbf{x}_{k/n} &= (x_{k/n,0}, \widetilde{\mathbf{x}}_{k/n}) = (x_{k/n,0}, x_{k/n,1}, \ldots, x_{k/n,m}) \\ &= \left(1, \frac{p_{k/n,1}}{p_{(k-1)/n,1}}, \ldots, \frac{p_{k/n,m}}{p_{(k-1)/n,m}}\right).\end{aligned}$$

Recall that $x_{k/n,0}$, the wealth relative of cash, is always *1*. Here, $\widetilde{\mathbf{x}}_{k/n} = (x_{k/n,1}, \ldots, x_{k/n,m})$ is just $\mathbf{x}_{k/n}$ with the first coordinate truncated (i.e., the wealth relative vector of the m stocks, cash excluded). Given these wealth relatives, the wealth achieved by $\widehat{\mathbf{b}}^{(n)}_{k/n}$ by time T is given by

$$\begin{aligned}\text{(A.19)} \qquad \widehat{W}^{(n)}_T &= W_0 \prod_{k=1}^{Tn} \widehat{\mathbf{b}}^{(n)\prime}_{(k-1)/n} \mathbf{x}_{k/n} \\ &= W_0 \prod_{k=1}^{Tn} \left(1 + (\widetilde{\mathbf{x}}_{k/n} - \mathbf{1})' \mathbf{A}\left(\mathbf{s}_{\lfloor (k-1)/n \rfloor}\right) \psi^{(n)}_{(k-1)/n}\right) \qquad \text{(by (5.2))},\end{aligned}$$

where $\mathbf{1}$ denotes the m-dimensional vector $(1, \ldots, 1)$. Now we add and subtract $\psi^{\dagger(n)}_{k/n}$ to get

$$\begin{aligned}\widehat{W}^{(n)}_T = W_0 \prod_{k=1}^{Tn} \Big(1 &+ (\widetilde{\mathbf{x}}_{k/n} - \mathbf{1})' \mathbf{A}\left(\mathbf{s}_{\lfloor (k-1)/n \rfloor}\right) \left(\psi^{(n)}_{(k-1)/n} - \psi^{\dagger(n)}_{(k-1)/n}\right) \\ &+ (\widetilde{\mathbf{x}}_{k/n} - \mathbf{1})' \mathbf{A}\left(\mathbf{s}_{\lfloor (k-1)/n \rfloor}\right) \psi^{\dagger(n)}_{(k-1)/n}\Big).\end{aligned}$$

But note that from universality conditions 1 and 5 and equation (A.18),

$$\begin{aligned}&\left\| (\widetilde{\mathbf{x}}_{k/n} - \mathbf{1})' \mathbf{A}\left(\mathbf{s}_{\lfloor (k-1)/n \rfloor}\right) \left(\psi^{(n)}_{(k-1)/n} - \psi^{\dagger(n)}_{(k-1)/n}\right) \right\| \\ &\quad \le \sqrt{m} L_p L^{1/2}_{\mathbf{A},d} \left\| \psi^{(n)}_{(k-1)/n} - \psi^{\dagger(n)}_{(k-1)/n} \right\| \\ &\quad = O_{gt}(n^{\epsilon/4}).\end{aligned}$$

Thus,

$$\begin{aligned}\text{(A.20)} \qquad \widehat{W}^{(n)}_T &\ge W_0 \prod_{k=1}^{Tn} \left(1 - O_{gt}(n^{\epsilon/4}) + (\widetilde{\mathbf{x}}_{k/n} - \mathbf{1})' \mathbf{A}\left(\mathbf{s}_{\lfloor (k-1)/n \rfloor}\right) \times \psi^{\dagger(n)}_{(k-1)/n}\right) \\ &= W_0 \prod_{k=1}^{Tn} \left(1 + \left(\widetilde{\mathbf{x}}_{k/n} - \mathbf{1}\right)' \mathbf{A}\left(\mathbf{s}_{\lfloor (k-1)/n \rfloor}\right) \psi^{\dagger(n)}_{(k-1)/n}\right) \\ &\quad \times \prod_{k=1}^{Tn} \left(1 - O_{gt}(n^{\epsilon/4}) \left(1 + (\widetilde{\mathbf{x}}_{k/n} - \mathbf{1})' \mathbf{A}\left(\mathbf{s}_{\lfloor (k-1)/n \rfloor}\right) \psi^{\dagger(n)}_{(k-1)/n}\right)^{-1}\right).\end{aligned}$$

We work with the two products in (A.20) separately. First note that for sufficiently large N, $\left\| (\widetilde{\mathbf{x}}_{k/n} - \mathbf{1})' \mathbf{A}\left(\mathbf{s}_{\lfloor (k-1)/n \rfloor + 1}\right) \psi^{\dagger(n)}_{(k-1)/n} \right\|$ will be bounded by some constant less than 1 for all $n > N$, and hence,

(A.21)
$$\begin{aligned}
&\prod_{k=1}^{Tn}\left(1 - O_{gt}(n^{\epsilon/4})(1+\left(\widetilde{\mathbf{x}}_{k/n}-\mathbf{1})'\mathbf{A}\left(\mathbf{s}_{\lfloor (k-1)/n\rfloor}\right)\psi^{\dagger(n)}_{(k-1)/n}\right)^{-1}\right)\\
&\quad=\prod_{k=1}^{Tn}\left(1 - O_{gt}(n^{\epsilon/4})\right)\\
&\quad=\left(1 - O_{gt}(n^{\epsilon/4})\right)^{Tn}\\
&\quad=\exp\left\{-Tn\,O_{gt}(n^{\epsilon/4})\right\}.
\end{aligned}$$

As for the other product, note that

$$\begin{aligned}
&W_0\prod_{k=1}^{Tn}\left(1+(\widetilde{\mathbf{x}}_{k/n}-\mathbf{1})'\mathbf{A}\left(\mathbf{s}_{\lfloor (k-1)/n\rfloor}\right)\psi^{\dagger(n)}_{(k-1)/n}\right)\\
&\quad= W_0\prod_{k=1}^{Tn}\left(1+(\widetilde{\mathbf{x}}_{k/n}-\mathbf{1})'\mathbf{A}\left(\mathbf{s}_{\lfloor (k-1)/n\rfloor}\right)\frac{\int_{\Theta_n}\boldsymbol{\theta}\,dG_{k-1,n}(\boldsymbol{\theta})}{\int_{\Theta_n}dG_{k-1,n}(\boldsymbol{\theta})}\right)\\
&\quad= W_0\prod_{k=1}^{Tn}\left(1+(\widetilde{\mathbf{x}}_{k/n}-\mathbf{1})'\mathbf{A}\left(\mathbf{s}_{\lfloor (k-1)/n\rfloor}\right)\frac{\int_{\Theta_n}\boldsymbol{\theta}\exp\left\{\mathbf{q}_{(k-1)/n}(\boldsymbol{\theta})\right\}d\pi(\boldsymbol{\theta})}{\int_{\Theta_n}\exp\left\{\mathbf{q}_{(k-1)/n}(\boldsymbol{\theta})\right\}d\pi(\boldsymbol{\theta})}\right)\\
&\quad= W_0\prod_{k=1}^{Tn}\left[\left(\int_{\Theta_n}\left(1+(\widetilde{\mathbf{x}}_{k/n}-\mathbf{1})'\mathbf{A}\left(\mathbf{s}_{\lfloor (k-1)/n\rfloor}\right)\boldsymbol{\theta}\right)\times\exp\left\{\mathbf{q}_{(k-1)/n}(\boldsymbol{\theta})\right\}d\pi(\boldsymbol{\theta})\right)\right.\\
&\qquad\qquad\left.\Big/\int_{\Theta_n}\exp\left\{\mathbf{q}_{(k-1)/n}(\boldsymbol{\theta})\right\}d\pi(\boldsymbol{\theta})\right].
\end{aligned}$$

But by Lemma 5.1 and the definition of $\mathbf{q}_{k,n}(\boldsymbol{\theta})$ recall that $W^{(n)}_{k/n}(\boldsymbol{\theta}) = \exp\{\mathbf{q}_{k,n}(\boldsymbol{\theta}) + \epsilon_{k/n}(\boldsymbol{\theta})\}$. Hence,

$$\begin{aligned}
&= W_0\prod_{k=1}^{Tn}\left[\left(\int_{\Theta_n}\left(1+(\widetilde{\mathbf{x}}_{k/n}-\mathbf{1})'\mathbf{A}\left(\mathbf{s}_{\lfloor (k-1)/n\rfloor}\right)\boldsymbol{\theta}\right)\right.\right.\\
&\qquad\left.\times W^{(n)}_{(k-1)/n}(\boldsymbol{\theta})\exp\left\{-\epsilon_{(k-1)/n}(\boldsymbol{\theta})\right\}d\pi(\boldsymbol{\theta})\right)\\
&\qquad\left.\Big/\int_{\Theta_n}W^{(n)}_{(k-1)/n}(\boldsymbol{\theta})\exp\left\{-\epsilon_{(k-1)/n}(\boldsymbol{\theta})\right\}d\pi(\boldsymbol{\theta})\right]\\
&= W_0\prod_{k=1}^{Tn}\left(\frac{\int_{\Theta_n}W^{(n)}_{k/n}(\boldsymbol{\theta})\exp\left\{-\epsilon_{(k-1)/n}(\boldsymbol{\theta})\right\}d\pi(\boldsymbol{\theta})}{\int_{\Theta_n}W^{(n)}_{(k-1)/n}(\boldsymbol{\theta})\exp\left\{-\epsilon_{(k-1)/n}(\boldsymbol{\theta})\right\}d\pi(\boldsymbol{\theta})}\right)\\
&= W_0\prod_{k=1}^{Tn}\left[\left(\int_{\Theta_n}W^{(n)}_{k/n}(\boldsymbol{\theta})\exp\{-\epsilon_{k/n}(\boldsymbol{\theta})\}\right.\right.\\
&\qquad\left.\times\exp\left\{\epsilon_{k/n}(\boldsymbol{\theta})-\epsilon_{(k-1)/n}(\boldsymbol{\theta})\right\}d\pi(\boldsymbol{\theta})\right)\\
&\qquad\left.\Big/\int_{\Theta_n}W^{(n)}_{(k-1)/n}(\boldsymbol{\theta})\exp\left\{-\epsilon_{(k-1)/n}(\boldsymbol{\theta})\right\}d\pi(\boldsymbol{\theta})\right].
\end{aligned}$$

We know from Lemma 5.1 that $|\epsilon_{k/n}(\boldsymbol{\theta}) - \epsilon_{(k-1)/n}(\boldsymbol{\theta})|$ is $O(B^3(1+\log n)^3/n^{3/2})$. Here B, the bound on $\|\mathbf{A}(\mathbf{s}_{\lfloor (k-1)/n\rfloor})\boldsymbol{\theta}\|_1$, grows at the same rate as Θ_n (i.e., $O(n^{\epsilon/4})$). Thus the absolute difference in error terms is $O\big((1+\log n)^3/n^{3/2-3\epsilon/4}\big)$. However this is in turn $O(n^{-3/2+\epsilon})$. Hence $|\epsilon_{k/n}(\boldsymbol{\theta}) - \epsilon_{(k-1)/n}(\boldsymbol{\theta})| = O(n^{-3/2+\epsilon})$ and

$$
\begin{aligned}
&\geq W_0 \prod_{k=1}^{Tn} \exp\{-O(n^{-3/2+\epsilon})\} \\
&\quad \times \left(\frac{\int_{\Theta_n} W_{k/n}^{(n)}(\boldsymbol{\theta}) \exp\{-\epsilon_{k/n}(\boldsymbol{\theta})\}\, d\pi(\boldsymbol{\theta})}{\int_{\Theta_n} W_{(k-1)/n}^{(n)}(\boldsymbol{\theta}) \exp\{-\epsilon_{(k-1)/n}(\boldsymbol{\theta})\}\, d\pi(\boldsymbol{\theta})} \right) \\
&= W_0 \exp\{-O(Tn^{-1/2+\epsilon})\} \\
&\quad \times \prod_{k=1}^{Tn} \left(\frac{\int_{\Theta_n} W_{k/n}^{(n)}(\boldsymbol{\theta}) \exp\{-\epsilon_{k/n}(\boldsymbol{\theta})\}\, d\pi(\boldsymbol{\theta})}{\int_{\Theta_n} W_{(k-1)/n}^{(n)}(\boldsymbol{\theta}) \exp\{-\epsilon_{(k-1)/n}(\boldsymbol{\theta})\}\, d\pi(\boldsymbol{\theta})} \right) \\
&= W_0 \exp\{-O(Tn^{-1/2+\epsilon})\} \\
&\quad \times \frac{\int_{\Theta_n} W_T^{(n)}(\boldsymbol{\theta}) \exp\{-\epsilon_T(\boldsymbol{\theta})\}\, d\pi(\boldsymbol{\theta})}{\int_{\Theta_n} W_0 d\pi(\boldsymbol{\theta})}.
\end{aligned}
$$

Again by Lemma 5.1 and the reasoning above, $|\epsilon_T(\boldsymbol{\theta})|$ is $O(Tn^{-1/2+\epsilon})$. Also, $\int_{\Theta_n} W_0\, d\pi(\boldsymbol{\theta}) = W_0\pi(\Theta_n) < W_0/C$ for some constant C, so the above is lower bounded by

$$
\geq C \exp\{-O(Tn^{-1/2+\epsilon})\} \int_{\Theta_n} W_T^{(n)}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta}). \tag{A.22}
$$

We now substitute the bounds (A.21) and (A.22) back into (A.20) to get a lower bound on $\widehat{W}_T^{(n)}$. However we should note that the $\exp\{-Tn\, O_{gt}(n^{\epsilon/4})\}$ bound of (A.21) is itself bounded by something of order $\exp\{-O(Tn^{-1/2+\epsilon})\}$ and thus the factor corresponding to (A.21) can be absorbed into the $\exp\{-O(Tn^{-1/2+\epsilon})\}$ factor of (A.22) and we conclude that

$$
\widehat{W}_T^{(n)} \geq C \exp\{-O(Tn^{-1/2+\epsilon})\} \int_{\Theta_n} W_T^{(n)}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta}). \tag{A.23}
$$

The next step in the proof is showing that $\int_{\Theta_n} W_T^{(n)}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})$ is close to $\widehat{W}_T$, the wealth achieved by the continuously traded universal portfolio $\widehat{\mathbf{b}}_t$ of equation (4.9). Recall by Lemma 4.1 that $\widehat{W}_T = \int_{\mathbf{R}^d} W_T(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})$. Since $W_T^{(n)}(\boldsymbol{\theta})$ converges to $W_T(\boldsymbol{\theta})$ as the number of rebalances increases and since Θ_n increases to all of $\mathbf{R}^d$ it makes intuitive sense that $\int_{\Theta_n} W_T^{(n)}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})$ converges to $\widehat{W}_T$ and we now show that this order of convergence is $\exp\{-O(Tn^{-1/2+\epsilon})\}$.

First examine the wealth ratio $W_T(\boldsymbol{\theta})/W_T^{(n)}(\boldsymbol{\theta})$. From the expressions of $W_T(\boldsymbol{\theta})$ and $W_T^{(n)}(\boldsymbol{\theta})$ in Theorem 4.1 and Lemma 5.1, we deduce that

$$
\begin{aligned}
\frac{W_T(\boldsymbol{\theta})}{W_T^{(n)}(\boldsymbol{\theta})} = \exp\Bigg\{ &\frac{1}{2} \sum_{\tau=1}^{T} \left(\operatorname{diag}\left(\mathbf{K}_\tau - \mathbf{K}_\tau^{(n)}\right)\right)' \mathbf{A}\left(\mathbf{s}_{\lceil \tau \rceil - 1}\right) \boldsymbol{\theta} \\
&- \sum_{\tau=1}^{T} \frac{1}{2} \boldsymbol{\theta} A'\left(\mathbf{s}_{\lceil \tau \rceil - 1}\right) \left(\mathbf{K}_\tau^{\dagger} - \mathbf{K}_\tau^{\dagger(n)}\right) \mathbf{A}\left(\mathbf{s}_{\lceil \tau \rceil - 1}\right) \boldsymbol{\theta} + \varepsilon_T(\boldsymbol{\theta}) \Bigg\}.
\end{aligned}
$$

Bounding the absolute value of the exponent, note by the universality conditions that

$$\left|\frac{1}{2}\sum_{\tau=1}^{T}\left(\operatorname{diag}\left(\mathbf{K}_{\tau}-\mathbf{K}_{\tau}^{(n)}\right)\right)'\mathbf{A}\left(\mathbf{s}_{\lceil\tau\rceil-1}\right)\boldsymbol{\theta}\right.$$

$$\left.-\sum_{\tau=1}^{T}\frac{1}{2}\boldsymbol{\theta}A'\left(\mathbf{s}_{\lceil\tau\rceil-1}\right)\left(\mathbf{K}_{\tau}^{\dagger}-\mathbf{K}_{\tau}^{\dagger(n)}\right)\mathbf{A}\left(\mathbf{s}_{\lceil\tau\rceil-1}\right)\boldsymbol{\theta}+\varepsilon_{k/n}(\boldsymbol{\theta})\right|$$

$$\leq\frac{1}{2}\sum_{\tau=1}^{T}\left\|\operatorname{diag}\left(\mathbf{K}_{\tau}-\mathbf{K}_{\tau}^{(n)}\right)\right\|\left\|\mathbf{A}\left(\mathbf{s}_{\lceil\tau\rceil-1}\right)\boldsymbol{\theta}\right\|$$

$$+\frac{1}{2}\sum_{\tau=1}^{T}|\lambda|_{\max}\left(\mathbf{K}_{\tau,j,j}^{\dagger}-\mathbf{K}_{\tau,j,j}^{\dagger(n)}\right)\left\|\mathbf{A}\left(\mathbf{s}_{\lceil\tau\rceil-1}\right)\boldsymbol{\theta}\right\|^{2}$$

$$\leq\frac{1}{2}L_{\mathbf{A},d}^{1/2}\|\boldsymbol{\theta}\|\sum_{\tau=1}^{T}\sqrt{m}\,|\lambda|_{\max}\left(\mathbf{K}_{\tau,j,j}^{\dagger}-\mathbf{K}_{\tau,j,j}^{\dagger(n)}\right)$$

$$+\frac{1}{2}L_{\mathbf{A},d}\|\boldsymbol{\theta}\|^{2}\sum_{\tau=1}^{T}|\lambda|_{\max}\left(\mathbf{K}_{\tau,j,j}^{\dagger}-\mathbf{K}_{\tau,j,j}^{\dagger(n)}\right)$$

$$\leq\frac{\sqrt{m}}{2}L_{\mathbf{A},d}^{1/2}\frac{L_k'}{\sqrt{n}}T\|\boldsymbol{\theta}\|+\frac{1}{2}L_{\mathbf{A},d}\frac{L_k'}{\sqrt{n}}T\|\boldsymbol{\theta}\|^{2}.$$

But since $\max_{\boldsymbol{\theta}\in\Theta_n}\|\boldsymbol{\theta}\|$ is $O(n^{\epsilon/4})$, it follows that the above is $O(Tn^{-1/2+\epsilon/2})$, which is in turn boundable by something $O(Tn^{-1/2+\epsilon})$, so

$$\frac{W_T(\boldsymbol{\theta})}{W_T^{(n)}(\boldsymbol{\theta})}\leq\exp\{O(Tn^{-1/2+\epsilon})\}.$$

Using this bound we note that

$$\int_{\Theta_n}W_T^{(n)}(\boldsymbol{\theta})\,d\pi(\boldsymbol{\theta})\geq\exp\{-O(Tn^{-1/2+\epsilon})\}\int_{\Theta_n}W_T(\boldsymbol{\theta})\,d\pi(\boldsymbol{\theta}),$$

and seeing that $\widehat{W}_T=\int_{\mathbf{R}^d}W_T(\boldsymbol{\theta})\,d\pi(\boldsymbol{\theta})$ we write

$$\int_{\Theta_n}W_T^{(n)}(\boldsymbol{\theta})\,d\pi(\boldsymbol{\theta})\geq\exp\{-O(Tn^{-1/2+\epsilon})\}\times\left(\int_{\Theta_n}\frac{W_T(\boldsymbol{\theta})\,d\pi(\boldsymbol{\theta})}{\int_{\mathbf{R}^d}W_T(\boldsymbol{\theta})\,d\pi(\boldsymbol{\theta})}\right)\widehat{W}_T. \tag{A.24}$$

Now define the measure

$$dG_T\equiv\frac{W_T(\boldsymbol{\theta})\,d\pi(\boldsymbol{\theta})}{\int_{\mathbf{R}^d}W_T(\boldsymbol{\theta})\,d\pi(\boldsymbol{\theta})}.$$

Using the same arguments used with $dG_{k,n}(\boldsymbol{\theta})$ of (A.12) we argue that G_T is Gaussian with mean ψ_T and covariance $\left(\boldsymbol{\Omega}_T^{-1}+\boldsymbol{\Lambda}^{-1}\right)^{-1}$. Increasing values of T only make the density more peaked and ψ_T is always within a bounded distance of the origin regardless of T. This along with the fact that Θ_n is getting larger with n is sufficient to argue that there exists a constant C independent of T and n such that

$$\int_{\Theta_n} \frac{W_T(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})}{\int_{\mathbf{R}^d} W_T(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})} = \int_{\Theta_n} dG_T > C.$$

Thus it follows from (A.24) that

$$\int_{\Theta_n} W_T^{(n)}(\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta}) \geq C \exp\{-O(Tn^{-1/2+\epsilon})\}\, \widehat{W}_T.$$

Substituting this bound into (A.23) we can increase C and merge the $\exp\{-O(Tn^{-1/2+\epsilon})\}$ terms in both the above inequality and (A.23), changing constants as necessary to deduce that

$$\widehat{W}_T^{(n)} \geq C \exp\{-O(Tn^{-1/2+\epsilon})\}\, \widehat{W}_T.$$

To complete the proof, recall from Theorem 4.2 that $W_T^* / \widehat{W}_T$ is $O(T^{d/2})$. Taking this and inverting the above inequality (and redefining C to be $1/C$), write

$$\begin{aligned} \frac{W_T^*}{\widehat{W}_T^{(n)}} &\leq C \exp\{O(Tn^{-1/2+\epsilon})\} \frac{W_T^*}{\widehat{W}_T} \\ &= C\, O(T^{d/2}) \exp\{O(Tn^{-1/2+\epsilon})\}. \end{aligned}$$

Redefining C as necessary to incorporate the constant associated with the $O(T^{d/2})$ term and defining α to be the constant associated with the $O(Tn^{-1/2+\epsilon})$ term, we conclude that

$$\frac{W_T^*}{\widehat{W}_T^{(n)}} \leq C T^{d/2} \exp\{\alpha T / n^{1/2-\epsilon}\}. \qquad \square$$

REFERENCES

COVER, T. M. (1991): Universal Portfolios, *Math. Finance* 1(1), 1–29.

COVER, T. M., and E. ORDENTLICH (1996): Universal Portfolios with Side Information, *IEEE Trans. Info. Theory* 42(2), 348–363.

GOLUB, G. H., and C. F. VAN LOAN (1997): *Matrix Computation*, 3rd ed. Baltimore: The Johns Hopkins University Press.

JAMSHIDIAN, F. (1992): Asymptotically Optimal Portfolios, *Math. Finance* 2(2), 131–150.

LARSON, D. C. (1986): *Growth Optimal Trading Strategies*. PhD thesis, Stanford Universit, Stanford, CA,

MERTON, R. C. (1969): Lifetime Portfolio Selection Under Uncertainty: The Continuous Time Case, *Rev. Econ. Stat.* 51, 247–257.

MERTON, R. C. (1971): Optimum Consumption and Portfolio Rules in a Continuous Time Model, *J. Econ. Theory* 3, 373–413.

SINGER, Y., D. P. HELMBOLD, R. E. SCHAPIRE, and M. K. WARMUTH (1998): On-Line Portfolio Selection Using Multiplicative Updates, *Math. Finance* 8(4), 325–347.